ESCUELA POLITÉCNICA NACIONAL

FACULTAD DE INGENIERÍA ELECTRÍCA Y ELECTRÓNICA

ANÁLISIS AUTOMATIZADO DE POLÍTICAS DE PRIVACIDAD EN ECUADOR

TRABAJO DE INTEGRACIÓN CURRICULAR PRESENTADO COMO REQUISITO PARA LA OBTENCIÓN DEL TÍTULO DE INGENIERO EN TECNOLOGÍAS DE LA INFORMACIÓN

JAIME PATRICIO RAMÍREZ RAMÍREZ

jaime.ramirez01@epn.edu.ec

DIRECTOR: JOSÉ ANTONIO ESTRADA JIMÉNEZ

jose.estrada@epn.edu.ec

DMQ, febrero 2022

CERTIFICACIONES

Yo, JAIME PATRICIO RAMÍREZ RAMÍREZ declaro que el trabajo de integración curricular aquí descrito es de mi autoría; que no ha sido previamente presentado para ningún grado o calificación profesional; y, que he consultado las referencias bibliográficas que se incluyen en este documento.

JAIME PATRICIO RAMÍREZ RAMÍREZ

Certifico que el presente trabajo de integración curricular fue desarrollado por JAIME PATRICIO RAMÍREZ RAMÍREZ, bajo mi supervisión.

JOSÉ ANTONIO ESTRADA JIMÉNEZ
DIRECTOR

DECLARACIÓN DE AUTORÍA

A través de la presente declaración, afirmamos que el trabajo de integración curricular aquí descrito, así como el (los) producto(s) resultante(s) del mismo, son públicos y estarán a disposición de la comunidad a través del repositorio institucional de la Escuela Politécnica Nacional; sin embargo, la titularidad de los derechos patrimoniales nos corresponde a los autores que hemos contribuido en el desarrollo del presente trabajo; observando para el efecto las disposiciones establecidas por el órgano competente en propiedad intelectual, la normativa interna y demás normas.

JAIME PATRICIO RAMÍREZ RAMÍREZ

JOSÉ ANTONIO ESTRADA JIMÉNEZ

DEDICATORIA

A mis padres, a mi hermana y al Dr. Harry, a quienes dedico todo mi esfuerzo y trabajo.

AGRADECIMIENTO

A mis profesores, particularmente a mi tutor el PhD. José Estrada Jiménez por toda la dedicación y enseñanzas brindadas. Al PhD. Gabriel López, quién me impulsó a continuar los estudios de mi carrera y al MSc. Pablo Hidalgo por todo su apoyo durante toda mi carrera universitaria. A mi amada Familia, especialmente a mis padres, a mi hermana y al Dr. Harry quienes han sido mi pilar fundamental. A mis amigos: Jair, Cynthia y Fabricio quienes me han apoyado incondicionalmente.

ÍNDICE DE CONTENIDO

RESUMEN

Ecuador implementó en 2021 la Ley Orgánica de Protección de Datos Personales (LOPDP), que debería cambiar la manera en que las instituciones tratan los datos personales de los usuarios. En ese contexto, las políticas de privacidad son una herramienta que permite garantizar el derecho de los usuarios a ser informados, ya que representan la intención que tienen las entidades para cumplir con las normativas legales.

Este trabajo presenta el análisis automatizado y manual de las políticas de privacidad de más de 50 sitios web ecuatorianos (escritas en español) presentes en el ranking Alexa. El análisis automatizado se fundamenta en una aproximación cuantitativa mediante el conteo de palabras y oraciones, cálculo de similitud entre políticas de privacidad, así como métodos de procesamiento de lenguaje natural que permita obtener la frecuencia de aparición de las palabras dentro de las políticas de privacidad, y el modelado de tópicos. Con el fin de contrastar los resultados obtenidos de este análisis, se realizó un análisis manual mediante encuestas junto a un análisis exploratorio de particularidades.

Los resultados obtenidos presentan un panorama preocupante para el país, debido a que, la extensión de las políticas de privacidad, su estructura no definida y falta de estandarización junto con la extrema generalización de su contenido, dificultan el acceso eficiente a su contenido por parte de los usuarios. Finalmente, se han descrito un conjunto de prácticas recomendadas que permitan a las entidades generar políticas de privacidad más amigables con el usuario, y que cumplan con la LOPDP.

PALABRAS CLAVE: políticas, privacidad, protección, datos personales, Ecuador.

ABSTRACT

Ecuador implemented in 2021 the Organic Law for the Protection of Personal Data (LOPDP), which should change the way in which institutions treat users' personal data. In this context, privacy policies are a tool that guarantees the right of users to be informed, since they represent the intention of entities to comply with legal regulations.

This work presents the automated and manual analysis of the privacy policies of more than 50 Ecuadorian websites (written in Spanish) present in the Alexa ranking. The automated analysis is based on a quantitative approach by counting words and sentences, calculating the similarity between privacy policies, as well as natural language processing methods to obtain the frequency of occurrence of words within the policies. privacy and modeling issues. To compare the results obtained from this analysis, a manual analysis was carried out using surveys together with an exploratory analysis of particularities.

The results obtained present a troubling panorama for the country, since the extension of privacy policies, their indefinite structure, and the lack of standardization together with the extreme generalization of their contents, make it difficult for users to have efficient access to their contents. Finally, a set of good practices have been described that allow entities to generate friendlier privacy policies and comply with the LOPDP.

KEYWORDS: policies, privacy, protection, personal data, Ecuador

1 INTRODUCCIÓN

La privacidad es un derecho humano proclamado en la Declaración Universal de Derechos Humanos [1] en 1984. Debido a la importancia de la privacidad para el desarrollo del ser humano, de hecho, la privacidad es la libertad más esencial [2]. A pesar de esto, es un derecho que no ha recibido la atención que debería [3], especialmente durante la cuarta revolución industrial que vivimos [4].

Actualmente, la sociedad globalizada provoca que la mayor parte de individuos estén conectados en todos los aspectos (social, cultural, político, etc.) usando medios tecnológicos [5]. La consecuencia de esto es que gran cantidad de información personal de todos los individuos se encuentra en la nube.

La disponibilidad de mucha información en la nube tiene grandes ventajas como: automatización en la toma de decisiones para personalizar servicios o productos, o facilitar trámites en instituciones estatales o privadas. La información es el recurso del futuro y, en ese sentido, permite incluso dotar de poder económico y político a quien la posee (por ejemplo, a un estado) [6].

Si bien las ventajas de tener grandes cantidades de información personal en la nube son muchas, existen desventajas que son preocupantes. Al estar tan disponible, esta información puede ser usada por delincuentes; asimismo, las entidades que la procesan pueden abusar de ella, por ejemplo, manipulando a sus dueños. Incluso su difusión puede causar un impacto negativo. La información sensible puede ser usada de varias maneras por las organizaciones o delincuentes que busquen hacer daño a los individuos sujetos de dicha información.

Debido a estos problemas, la privacidad y protección de datos personales debe ser un pilar fundamental en las labores del estado. En esa línea, en mayo del 2021, se aprobó en Ecuador la Ley Orgánica de Protección de Datos Personales (LOPDP) [7] buscando tener una base legal que dote al estado de herramientas para precautelar el derecho a la privacidad. Aunque la LOPDP es una herramienta regulatoria poderosa, su aplicación debe ser guiada por un diagnóstico que permita determinar el estado actual de la privacidad en el país, y que puede tener distintas aristas.

Tal como se ha verificado en trabajos previos [8], una de esas aristas consiste en comprender la calidad de la protección de datos personales aplicada por las instituciones, por ejemplo, mediante el análisis de las políticas de privacidad que estas publican. Esto es posible debido a que las políticas de privacidad son documentos formales que ofrecen una

declaración de intenciones respecto a la protección de datos personales y también proveen una idea general del actuar de las instituciones frente a estos datos. Su estudio, de hecho, sería útil para analizar la situación antes y después de la implementación obligatoria de la LOPDP.

Como un aporte en ese análisis, el presente trabajo busca analizar el contenido de las políticas de privacidad de 67 sitios web populares de Ecuador. Este análisis sirve de base para medir la evolución de estos documentos y verificar las mejoras que estas presenten, una vez que la ley se haya aplicado. Adicionalmente, este trabajo plantea comparar las políticas de privacidad en Ecuador con las de 48 sitios web de México, un país que cuenta con una ley de protección de Datos Personales desde julio de 2010 [9], por lo que seguramente su aplicación está mucho más difundida que en Ecuador.

El análisis de las políticas de privacidad se fundamenta principalmente en el análisis cuantitativo de las palabras y oraciones de cada política, así como en el nivel de similitud entre estas. Adicionalmente, se verifica la prevalencia de los diferentes términos, junto con la detección de tópicos, lo cual permite verificar los diferentes temas tratados en ellas.

Adicionalmente, se analizan encuestas realizadas a usuarios sobre la facilidad para entender y leer políticas de privacidad, y sobre el cumplimiento de estas con el derecho a la información del usuario ofrecido en la ley.

1.1 OBJETIVO GENERAL

Analizar de forma automatizada políticas de privacidad en Ecuador

1.2 OBJETIVOS ESPECÍFICOS

1. Identificar prácticas recomendadas para la elaboración de políticas de privacidad.
2. Generar una base de datos de políticas de privacidad de sitios web ecuatorianos y de otro país latinoamericano.
3. Analizar el texto de las políticas de privacidad para su caracterización.
4. Evaluar las características de las políticas de privacidad frente a las prácticas recomendadas.

1.3 ALCANCE

El trabajo de titulación propuesto se enfoca en el análisis de políticas de privacidad en Ecuador y México. Con ese fin, se revisará inicialmente las normativas local e internacional

relacionadas con privacidad y protección de datos personales, que sirvan de guía para la elaboración de políticas de privacidad. Esto permitirá tener una visión de la estructura esperada de este documento.

En esta línea, se buscará en la literatura científica trabajos vinculados con el estudio de políticas de privacidad o documentos similares con el fin de determinar las herramientas y las metodologías más adecuadas para su análisis tanto manual como automatizado.

Para la recolección de datos, primeramente, se seleccionará un grupo de al menos 50 sitios web ecuatorianos que sea muy populares en el ámbito nacional o que, por el servicio o contenido que ofrezcan, sea más probable que recolecten información de sus usuarios. Estas características incrementarían la necesidad de disponer de una política privacidad conforme a los mandatos legales. Esta recolección y otros pasos posteriores se ejecutarán también para México, específicamente uno que tenga legislación de protección de datos más madura, y que permita así un contraste con el contexto ecuatoriano.

Con fines de análisis posterior, la ubicación en la Web del documento de política de privacidad será registrada. Este enlace de cada política de privacidad, de existir, se registrará con una marca de tiempo, y se archivará también en línea, en un servicio como Internet Archive. Además, este documento será almacenado localmente en formato de texto claro, en dos versiones: una, tal como esté publicado y otra procesada, solamente incluyendo el texto pertinente. Estos pasos permitirán la construcción de un conjunto de datos (*data set*) de políticas de privacidad, que permitirá el análisis automatizado posterior. Para este proceso de recolección, se emplearán librerías basadas en Python.

Una vez esté disponible el conjunto de datos, se procesará con mecanismos automatizados el texto de las políticas de privacidad. Este proceso considerará técnicas de procesamiento de lenguaje natural como conteo de palabras, identificación de tópicos clave, y frecuencia con la cual aparecen las palabras.

Ya que este análisis buscará contrastar el estado de la privacidad localmente con el de otro país, también se analizarán manualmente las políticas de privacidad, para identificar los principales componentes estructurales que no se puedan obtener automáticamente.

El registro de la ubicación de las políticas de privacidad permitirá la medición de la prevalencia de estas políticas en sitios web en Ecuador. Este podría ser un indicador clave que determine la evolución de la implementación de la Ley de Protección de Datos en el tiempo.

A partir del estudio de esta normativa y de la recolección y análisis de los datos, se verificará la concordancia de las políticas de privacidad con la regulación local, esto es la Ley Orgánica de Protección de Datos Personales.

Finalmente, con todos estos insumos, se discutirá sobre el nivel de protección de datos personales en Ecuador y se ofrecerán recomendaciones para que este nivel se incremente en el marco de la normativa vigente.

Este trabajo no tendrá un producto final demostrable.

1.4 MARCO TEÓRICO

En esta sección se detallará inicialmente definiciones que son de utilidad para comprender el trabajo desarrollado. Con base en estas definiciones, luego, se analizará la base legal que garantiza el ejercicio del derecho a la privacidad en el país. Posteriormente, se detallará la información necesaria sobre políticas de privacidad que permitirá tener una visión ampliada de estas y su importancia. Finalmente, se realizará una revisión teórica de procesamiento de datos, y particularmente de las técnicas y herramientas utilizadas para el presente trabajo.

1.4.1 DEFINICIONES

A continuación, se presentan algunas definiciones importantes para la compresión de los conceptos abordados en este trabajo. La mayor parte de estas han sido obtenidas del capítulo I, artículo 4 de la LOPDP.

El primer concepto importante es el de *dato personal* que, de acuerdo con la Comisión Europea [10], es toda información perteneciente a una persona que permita identificarla de forma parcial o total. Ejemplos de datos personales son: nombre y apellidos, domicilio, documento nacional de identidad, dirección de protocolo de internet (IP), entre otros.

Por otro lado, la *privacidad* según la RAE [11], se define como el ámbito del desarrollo humano de la vida privada que el individuo tiene derecho a proteger de cualquier intromisión.

Otro término importante para destacar es el de *dato sensible,* y es todo dato cuyo almacenamiento y tratamiento debe ser especialmente protegido, pues su mal utilización podría causar, por ejemplo, discriminación. Algunos ejemplos incluyen: religión, raza, vida sexual, ideología política [12].

Un término importante es *titular* que, según la LOPDP [7], es la persona de la cual se procesará su información.

Ya que el titular es la persona que genera los datos, es necesario conocer quién es el *encargado del tratamiento de datos personales*. La LOPDP describe al *encargado* como la entidad que procesa los datos en nombre del responsable del tratamiento. El *responsable de tratamiento de datos personales*, en cambio, es aquella entidad que decide como se tratarán los datos y qué se hará con la información obtenida, según la LOPDP.

Una vez que se ha obtenido la información, se procede al *tratamiento* que es establecido en la LOPDP como la operación u operaciones automáticas o manuales que se realicen sobre los datos personales.

Otro concepto importante que se explicará en la sección 1.4.2.2, es el del *derecho a la información* que en el capítulo III de la LOPDP, se describe como el derecho del titular de los datos a ser informado de la finalidad de la recopilación de esos datos, del tratamiento que se les dará, con qué fin se realiza esto, y quién es el encargado de realizarlo.

Otro término interesante que es necesario definir, por su uso para la recopilación de datos es el de *cookie*. El diccionario panhispánico de español jurídico [12] lo define como aquel mecanismo utilizado por las páginas web para mejorar su usabilidad mediante la creación de un fichero en el computador del usuario, en el cual se almacenará información del usuario, así como sus gustos e intereses para que en el siguiente acceso se ofrezca una experiencia personalizada.

1.4.2 BASE LEGAL

1.4.2.1 Privacidad como Derecho Humano

El artículo 12 de la Carta de los Derechos Humanos establece que “Nadie será objeto de injerencias arbitrarias en su vida privada, su familia, su domicilio o su correspondencia, ni de ataques a su honra o a su reputación. Toda persona tiene derecho a la protección de la ley contra tales injerencias o ataques” [1]. La privacidad, desde esta declaración de los derechos humanos, se concibe como algo primordial y necesario en la vida de cada ser humano. La privacidad permite a los seres humanos auto descubrirse, conocerse mejor y se constituye en un pilar fundamental en el desarrollo de la personalidad. La identidad de cada individuo se forja en la intimidad ya que permite a los individuos actuar con normalidad y sin ningún condicionamiento de quien pueda estar observando; el deseo natural de evitar

ser observado constantemente se relaciona íntimamente con la sensación de estar protegidos, siendo esta una de las mayores aspiraciones humanas [2].

1.4.2.2 Privacidad en el Ámbito Jurídico Ecuatoriano

Países europeos como España cuentan con una ley de protección de datos personales consolidada desde 1999, denominada Ley Orgánica de Protección de Datos de Carácter Personal [13]. Actualmente esta ley y la del resto de países de la Unión Europea, fueron reemplazadas por el Reglamento General de Protección de Datos (RGPD) que está vigente desde mayo del 2016 [14]. La RGDP es una de las mejores y más completas leyes de protección de datos. Asimismo, Estados Unidos cuenta desde 1994 [15] con una gran cantidad de leyes específicamente orientadas a la protección de datos, pero no tiene una ley de protección de datos personales que regule el ámbito general. Por otro lado, Latinoamérica se ha visto rezagada en el desarrollo de normativa de protección de datos. México es uno de los países con la mejor legislación de protección de datos en la región, denominada Ley Federal de Protección de Datos Personales en Posesión de los Particulares, que se encuentra vigente desde 2010.

En Ecuador, en el ámbito regulatorio, la privacidad ha sido relegada a un segundo plano. Aunque el derecho a la privacidad, sí se mencionaba en algunas leyes, es apenas hace unos meses que se promulgó una Ley de Protección de Datos.

La Ley de Comercio Electrónico, Firmas y Mensajes de Datos [16] publicada en el Registro Oficial en abril de 2002 establece los primeros pasos del Estado para proteger la privacidad de los ciudadanos. Esta ley establece en la sección primera, artículo 9, que la recopilación y uso de datos personales se realizará respetando el derecho a la privacidad, intimidad y confidencialidad estipulados en la Constitución de la República del Ecuador del 1998 [17]. Esta ley establece el derecho a revocar el permiso para el tratamiento de datos y que el uso de datos de fuentes públicas es libre para la Administración Pública. Si bien esto es importante en cuestiones de privacidad, no se establecen los derechos básicos del titular de la información, como son el acceso a su información o el de rectificación de dicha información.

Posteriormente, en la Ley Orgánica de Telecomunicaciones [18] suscrita en el Registro Oficial el 18 de febrero de 2015, establece en el capítulo I, artículo 22, la necesidad de proteger los datos de los usuarios por parte de los prestadores de servicio, aplicando la normativa vigente. Sin embargo, no establece los mecanismos para conseguir esta protección. Esta ley es un gran paso en la generación de una normativa de protección de

datos personales ya que establece como un derecho de los usuarios y como una obligación de los prestadores de servicios la protección de los datos personales de los usuarios.

Finalmente, la Ley Orgánica de Protección de Datos Personales (LOPDP) publicada en el registro oficial el 26 de mayo de 2021, establece un marco normativo completo en el que se delinean los derechos de los ciudadanos sobre sus datos personales y cómo todas las instituciones deben respetar y proteger el cumplimiento de esta ley.

La LOPDP, al ser una ley general para la protección de datos personales, en su capítulo III incluye los derechos de los individuos en el marco de la protección de datos en Ecuador.

El artículo 12 de la LOPDP, establece el derecho a la información, basado en el principio de transparencia, por el cual el titular de los datos tiene derecho a conocer: los fines del tratamiento que se hará a sus datos, la base legal que se usará para su tratamiento, el tipo de tratamiento que se realizará (sea automático o manual), el tiempo que se almacenarán sus datos personales, la existencia de bases de datos que almacenen sus datos personales, el origen de los datos personales en caso de que el titular no sea quien provee dichos datos, el medio para contactar al responsable del tratamiento de datos personales y su identidad, las transferencias nacionales o internacionales de los datos personales, la consecuencia de entregar o no sus datos personales, las consecuencias de entregar información falsa o incorrecta, la posibilidad de revocar el consentimiento para tratar sus datos personales, la manera de aplicar los derechos al acceso, eliminación y portabilidad. Además, en caso de existir un delegado para la protección de datos, el titular tiene derecho a ser informado sobre el contacto e identidad de ese delegado, a conocer sobre la Autoridad de Protección de Datos Personales y, finalmente, si se está aplicando tratamiento de datos automatizado con perfilamiento.

Según el artículo 12 de la LOPDP, el individuo tiene derecho a conocer cómo ejercer su derecho de acceso a la información. Este derecho implica que el usuario pueda obtener de manera gratuita todos sus datos personales e información descrita en el párrafo anterior. Esta información debe ser entregada por el responsable del tratamiento sin solicitar una justificación y facilitando el acceso a este derecho a los usuarios en un plazo máximo de 15 días contados a partir de la fecha de petición realizada por el usuario.

Al igual que el derecho al acceso, el titular tiene derecho a la actualización o rectificación en caso de que sus datos personales sean erróneos o inexactos; para esto debe presentar la justificación debida al encargado del tratamiento de los datos, y este último debe realizar la actualización en máximo 15 días contados a partir de la fecha de petición realizada por el usuario.

El usuario puede solicitar la eliminación de sus datos personales, haciendo uso del derecho a eliminación. Esta eliminación puede ser solicitada cuando se haya cumplido la finalidad para la cual se recopilaron esos datos; cuando el plazo para su almacenamiento haya vencido, o cuando no se cumpla la ley o la finalidad de su recolección.

El usuario puede negarse a que se realice el tratamiento de sus datos desde el momento de su recolección; para esto, el derecho a la oposición permite al usuario oponerse al tratamiento de sus datos: siempre y cuando no se afecten los derechos de terceros, cuando el tratamiento se realiza con fines de mercadotecnia, o cuando no se ha solicitado el consentimiento para el tratamiento de los datos en los casos que este tratamiento sea requerido (por ejemplo, en temas relacionados con contratos).

El usuario puede exigir su derecho a la suspensión del tratamiento de sus datos al responsable de dicho tratamiento en los casos en que: los datos no sean correctos, de que el tratamiento de sus datos sea ilícito, de que el responsable haya cumplido con la finalidad del tratamiento, o que el usuario se niegue al tratamiento de sus datos personales médicos.

El usuario tiene varios derechos que le asisten para que sus datos no sean tratados, pero también es posible que el usuario desee transferir sus datos personales a otro responsable de tratamiento, para lo cual puede aplicar el derecho a la portabilidad, debiéndose verificar que los dos responsables de tratamiento (transmisor y receptor de los datos personales) cumplan con los lineamientos legales.

Como lo garantiza el derecho a ser informado, los titulares pueden conocer cuando sus datos son tratados por métodos automatizados que asisten en la toma de decisiones automatizadas. En caso de que el titular así lo desee, puede oponerse al tratamiento automatizado, haciendo uso de su derecho a no ser objeto de una decisión basada única o parcialmente en valoraciones automatizadas. Este derecho se extiende indicando que no se tratará datos sensibles de niños, niñas o adolescentes, a menos que se cuente con una autorización explícita del titular legal de dichos menores.

Una vez se han establecido los derechos básicos de los titulares respecto a la protección de datos personales, se establecen varios derechos que promueven la difusión de los conceptos relacionados con esta ley, con el objeto de evitar la vulneración de los derechos aquí descritos, ya sea por desconocimiento o falta de acceso a la información. En el marco de estos derechos, se encuentra el derecho a consulta. Este derecho, establece que todas las personas podrán consultar sus datos personales de manera gratuita en el Registro Nacional de Protección de Datos Personales. En este contexto se ha establecido también el derecho a la educación digital (el uso y manejo seguro, adecuado y responsable de las

tecnologías de la información y comunicación), para todas las personas, en todos los niveles de educación y en todos los campos de la educación.

1.4.3 POLÍTICA DE PRIVACIDAD

Una política de privacidad es un documento que contiene una declaración de intenciones sobre cómo la entidad o encargado del tratamiento de datos: recolectará, almacenará, protegerá y utilizará los datos personales del usuario. La política de privacidad es un documento formal en formato físico, electrónico, visual o sonoro, generado por el encargado del tratamiento. Debe ser otorgado o comunicado de manera oportuna al titular, para que este sea informado y en caso de ser necesario, acepte la recopilación y tratamiento de sus datos personales [19].

En el ámbito ecuatoriano, la política de privacidad es un documento que permite al titular hacer uso de su derecho a ser informado. En esa línea, la política de privacidad de la entidad debería contener toda la información resaltada previamente en el artículo 12 de la LOPDP.

Una buena política de privacidad según Website Policies [20] debe informar, adicionalmente a lo descrito en el artículo 12 de la LOPDP, sobre el uso de cookies para la recolección de la información y debe proveer un enlace para obtener la política de cookies o incluirla dentro de la política de privacidad. Adicionalmente, debido a que las políticas deben ser entregadas oportunamente a los titulares, se recomienda que el acceso a las políticas de privacidad sea sencillo. Por ejemplo, el acceso a la política debería ser posible desde la página principal del sitio web de la entidad, y esta debería encontrarse en una sección visible y fácil de encontrar.

En Internet se cuenta con un gran número de plantillas o guías para la generación de políticas de privacidad, las cuales se fundamentan en el artículo 12 del RGPD [21]. A partir de estas plantillas, las principales recomendaciones para la elaboración de políticas de privacidad serían:

- Toda política de privacidad debe ser precisa, por lo cual en ella se deberá incluir únicamente aquello que sea pertinente a la política, sin contener distractores que dificulten al titular su comprensión.
- De igual manera, para facilitar la comprensión de la política de privacidad, esta debe ser escrita en un lenguaje sencillo y claro, sin términos jurídicos que puedan causar

confusión en el titular, buscando siempre ser inteligible, y que pueda ser comprendida por todas las personas.

- Al ser una política de privacidad un documento legal que soporta el cumplimiento de una ley debe contar con transparencia en todo el proceso de su acceso, siendo necesario un acceso sencillo y gratuito.

1.4.4 PROCESAMIENTO DE DATOS

A continuación, se presentarán los fundamentos teóricos del procesamiento de datos utilizado para el análisis de las políticas de privacidad. Esto incluye una descripción inicial del método utilizado para la obtención de las políticas de privacidad y, posteriormente, una explicación de la rama del procesamiento de datos utilizada en el presente trabajo.

1.4.4.1 Web Scraping

Web Scraping es un término anglosajón utilizado para referirse a la extracción y almacenamiento automatizado de datos provenientes de páginas web. Entre los datos que pueden ser extraídos utilizando *web scraping* se encuentran: números telefónicos, correos electrónicos, enlaces, bases de datos, documentos adjuntos, entre otros [22].

El proceso de *web scraping* automatizado se fundamenta en el uso de *bots* para realizar la extracción de contenido web. El pilar fundamental de la técnica de *web scraping* es el análisis sintáctico, que se realiza a nivel de código HTML de una página web. El proceso de *web scraping* incluye actividades como el análisis del código para determinar qué información es la que se requiere almacenar; asimismo, en caso de que sea necesaria una navegación más profunda en el sitio web, esta se haría mediante el acceso a diferentes enlaces embebidos hasta localizar la información requerida.

Este método de recopilación de información web fue utilizado en el presente trabajo con el objeto de obtener las políticas de privacidad presentes en 67 sitios web del Ecuador de manera automatizada. Esto nos permitió comprobar, hasta cierto punto, la facilidad de acceso a las políticas de privacidad.

1.4.4.2 Procesamiento del Lenguaje Natural

El análisis automatizado de las políticas de privacidad que se realiza en este trabajo se fundamenta principalmente en el Procesamiento del Lenguaje Natural (PLN). Esta es una rama del procesamiento de datos orientada a la interacción en lenguaje humano entre un computador y el usuario (ej. para implementar *chat bots*). Esto es posible, ya que el PLN

permite a un computador analizar “texto claro” que no se encuentre estructurado de ninguna manera en particular.

Parte del análisis de las políticas realizado en este trabajo se fundamenta en tres aplicaciones de PLN: análisis de prevalencia de palabras y oraciones, análisis de similitud, y detección de tópicos relevantes.

Ya que cualquier documento que contiene lenguaje natural puede incluir información no relevante para el análisis realizado, especialmente si se trabaja con aprendizaje automático, usualmente es necesaria una etapa de preprocesamiento de los datos. Este preprocesamiento consiste en eliminar aquellas palabras que no aportan información relevante [23], por ejemplo: artículos, pronombres, preposiciones, entre otros. Adicionalmente, es necesario eliminar caracteres especiales y tildes para facilitar la generación de modelos de aprendizaje.

Tras finalizar el preprocesamiento se genera un “*token*” numérico que representa a cada palabra, permitiendo así al algoritmo de PLN trabajar más fácilmente con los datos. Una vez que se cuenta con *tokens* generados para cada palabra, determinar la frecuencia de estos, y así el número de palabras, resulta sencillo. La posibilidad de tener conjuntos de *tokens* correspondientes a cada política de privacidad permite aplicar fórmulas algebraicas sobre los datos; por ejemplo, para calcular la similitud coseno, que permite medir la similitud entre dos vectores [24]. Finalmente, el uso de *tokens* permite aplicar modelos ya generados como *Latent Dirichlet Allocation* (LDA), que permite detectar conjuntos de palabras relevantes para hallar tópicos en un documento.

El modelo LDA, se fundamenta en asignar pesos a diferentes tópicos, permitiendo así detectar la existencia de varios tópicos (temas) en un documento de texto. Esto permite relacionar una misma palabra en varios tópicos, si estas tienen el peso suficiente para ser consideradas relevantes dentro de dicho tópico. Este modelo se fundamenta en la Distribución de Dirichlet que describe la probabilidad de la pertenencia de un número a un conjunto dado.

1.4.5 TRABAJOS RELACIONADOS

1.4.5.1 Privacy Policies over Time: Curation and Analysis of a Million-Document Dataset

Este estudio se fundamenta en el análisis de más de un millón de políticas de privacidad escritas en idioma inglés, presentes en el top 100 del ranking Alexa de Estados Unidos y la Unión Europea [8].

1.4.5.2 The Creation and Analysis of a Website Privacy Policy Corpus

Este trabajo de investigación al igual que el anterior, se fundamenta en la generación de una base de datos de políticas de privacidad escritas en idioma inglés y presentes en el top 100 de sitios web más visitados en países angloparlantes [25].

Las referencias bibliográficas analizadas permiten verificar que el análisis de políticas de privacidad es inexistente en países hispanos, debido a lo cual el presente trabajo se centra en analizar políticas de privacidad escritas en castellano, presentes en el ranking Alexa de Ecuador que es una muestra de una comunidad muy particular como lo es Latinoamérica.

2 METODOLOGÍA

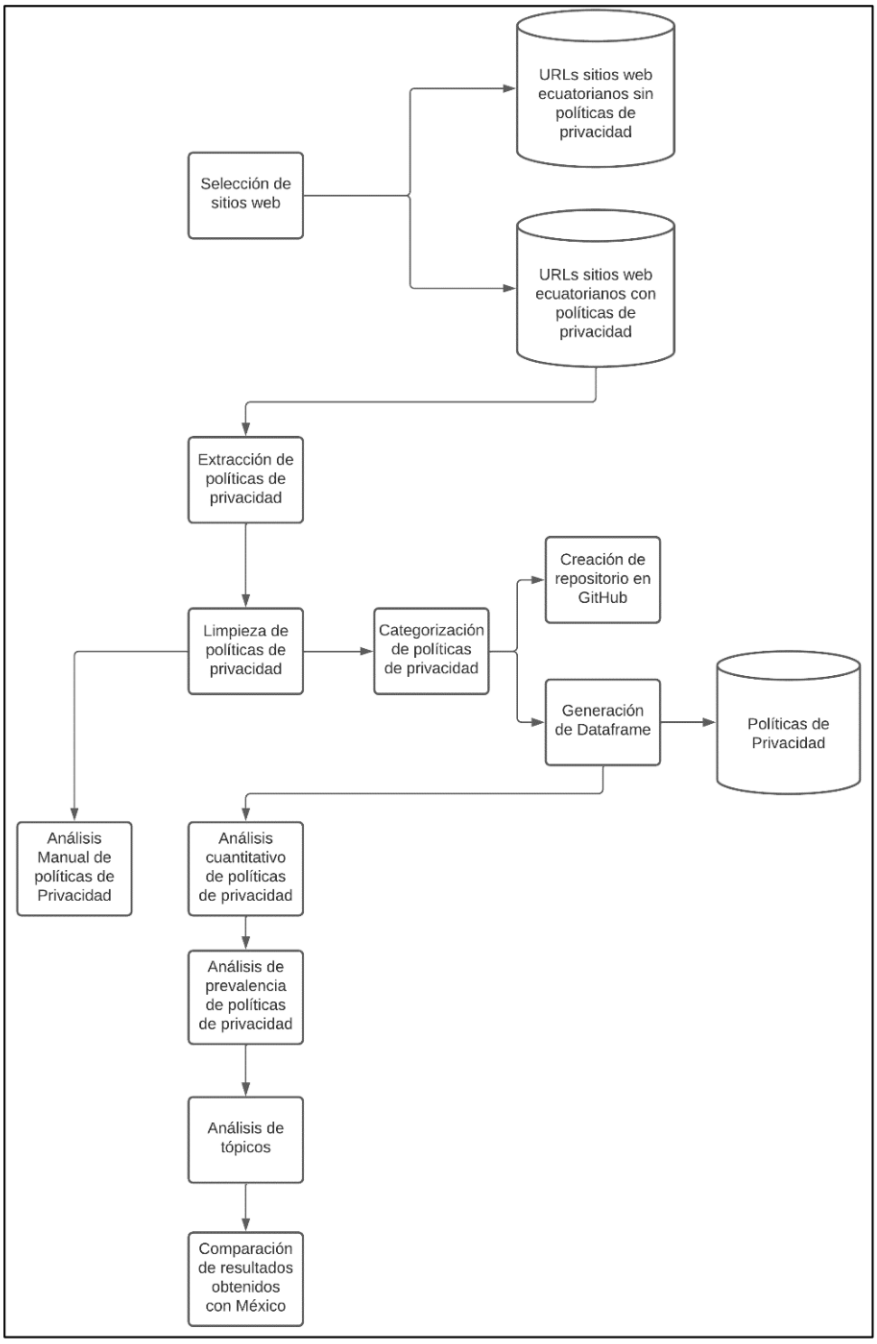


Figura 2.1. Esquema de Metodología Aplicada

2.1 GENERACIÓN DE BASE DE DATOS DE ENLACES DE POLÍTICAS DE PRIVACIDAD

Como punto de partida para la obtención de una muestra de políticas de privacidad, tanto de Ecuador como de México, se obtuvieron los enlaces de los sitios web locales más populares de ambos países. Para ello, se utilizó el ranking Alexa [26], que recopila los sitios

web más visitados mundialmente. Estos sitios tendrán, con mayor probabilidad, una política de privacidad. Ya que este estudio es una aproximación al análisis de la privacidad en Ecuador, se descartaron sitios de alcance internacional, por ejemplo, los de Facebook o Google.

Una vez que se ha determinado qué sitios web serán analizados, se identifica manualmente aquellos que cuenten con una política de privacidad. Así, se tienen dos bases de datos: una con los sitios web que poseen políticas de privacidad, y otra con los que no poseen políticas de privacidad.

Debido a que la aplicación de la LOPDP está atravesando un período de transición, es posible que los sitios web en Ecuador vayan ajustando sus políticas de privacidad en los próximos meses para cumplir dicha legislación. Para tener una imagen de esas políticas antes de que se modifiquen y medir dicha evolución, se procedió a almacenarlas en la plataforma Internet Archive: Wayback Machine [27].

Internet Archive, es una plataforma web que brinda el servicio de una librería digital de sitios web, así como contenido cultural (audios, libros, videos, software e imágenes). La plataforma fue creada por la organización sin fines de lucro Archive-IT [28], debido a que el objetivo de esta iniciativa es contar con la librería digital más grande permite a los usuarios almacenar archivos gratuitamente mientras estos se mantengan disponibles para todos los usuarios. Por tal motivo, resulta una plataforma idónea para almacenar los sitios web de las políticas de privacidad analizadas en el presente trabajo.

Para registrar cada sitio web en la plataforma, se siguió el siguiente procedimiento con los enlaces de las políticas de privacidad:

1. Acceder al servicio Internet Archive: Wayback Machine mediante el enlace https://web.archive.org/save/ como se muestra en la Figura 2.2.

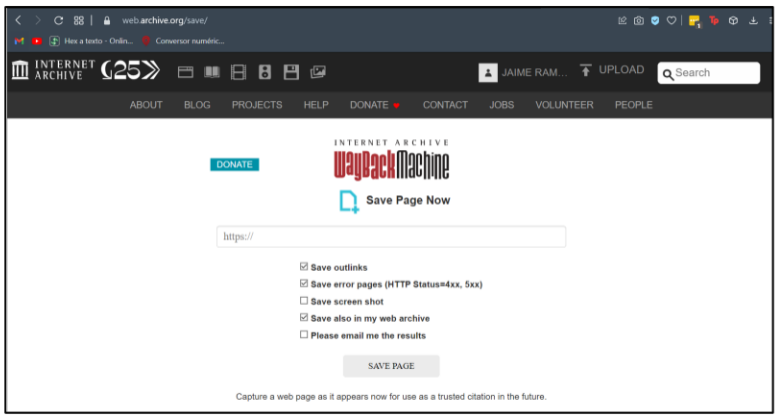


Figura 2.2. Plataforma Wayback Machine

2. Ingresar el enlace de la política de privacidad que se desea almacenar. A continuación, presionar el botón “SAVE PAGE”.

3. Una vez almacenado el sitio web, es posible apreciar el banner “Done!” que indica que, en este caso, la política de privacidad ha sido almacenada.

Figura 2.3. Sitio web con la política de privacidad almacenado

4. Finalmente, en la biblioteca principal de sitios web es posible apreciar las políticas de privacidad recopiladas, permitiendo tener un acceso al histórico de estas.

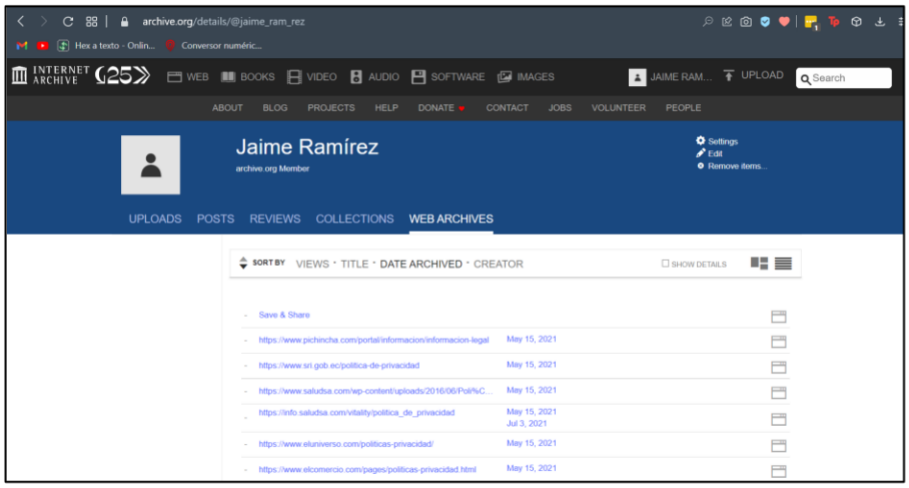


Figura 2.4. Biblioteca con histórico de políticas de privacidad.

El histórico con las políticas de privacidad recopiladas ha sido configurado de manera pública para facilitar el acceso a estas en futuros trabajos, su acceso es mediante el link https://archive.org/details/@jaime_ram_rez?tab=web-archive.

Una vez que se cuenta con un respaldo en la nube de las políticas de privacidad, se procede a generar una base de datos con formato .csv para facilitar su uso en posteriores procesos del presente trabajo.

La base de datos contiene las siguientes columnas: *Entity*, *Product*, *Entity URL Full*, *Entity URL*, *Policy URL*, *Comments*, *WayBack Machine URL*, *Category* y *Registration date*. A continuación, se detalla la información presente en cada campo, así:

- *Entity*: contiene la información correspondiente al nombre completo de la entidad a la cual pertenece la política de privacidad.
- *Product:* contiene el nombre del producto al cual hace referencia la política de privacidad. Una misma entidad puede contener varias políticas de privacidad correspondientes a diferentes productos o servicios.
- *Entity URL Full:* corresponde al URL del sitio web principal de la entidad, por ejemplo, https://www.sri.gob.ec/web/intersri/home. Este campo será necesario posteriormente para el proceso automatizado de descarga de políticas de privacidad.
- *Entity URL:* a diferencia del campo anterior, este campo contiene el URL reducido de la entidad, por ejemplo, sri.gob.ec. Este campo es usado en un futuro para etiquetar las políticas de privacidad, usado como nombre para cada política o como identificador de cada política durante su procesamiento.
- *Policy URL:* contiene el URL directo a la política de privacidad, para descargarla localmente.
- *Comments*: permite añadir observaciones para registrar algún inconveniente; por ejemplo: “No es posible almacenar en *Web Archive*”, o “Actualmente no disponible”. Estos mensajes se registrarán cuando la política no pueda ser almacenada debido a configuraciones de seguridad de ciertos sitios web, o cuando la política ya no esté disponible mediante el enlace registrado, respectivamente.
- *WayBack Machine URL*: es el enlace directo para acceder al histórico de la política de privacidad en la plataforma Wayback Machine.
- *Category*: contiene una categoría que se ha asignado manualmente a la entidad cuya política se registra, en función de su actividad, o el servicio/producto que ofrece. Esta etiqueta nos permite luego comprobar la similitud entre las políticas de privacidad de distintas entidades cuando tienen giros de negocio. Esta información

también podría ser usada para comparar las características de las políticas de privacidad en función de la categoría a la que pertenece el sitio web. Las categorías utilizadas en este estudio son: *education, employment, entertainment, finances, government, health, lottery, news, shopping, other*.

- *Registration date*: contiene la fecha en la cual se registró el URL de la política de privacidad en la base de datos.

2.2 EXTRACCIÓN DE POLÍTICAS DE PRIVACIDAD

Como se indicó antes en la sección 1.4.4.1, la extracción de las políticas de privacidad de los distintos sitios web se realiza mediante *web scraping,* que permite recopilar información de forma automatizada desde páginas web. En esta sección se detallará el proceso para localizar y extraer las políticas de privacidad a partir de los enlaces a las páginas web almacenadas en la base de datos con formato .csv, que fue generada en la sección anterior. Para esto se han desarrollado varios *scripts* en lenguaje Python 3.

A continuación, en la Figura 2.5 se presenta el diagrama del proceso usado para la extracción de políticas de privacidad.

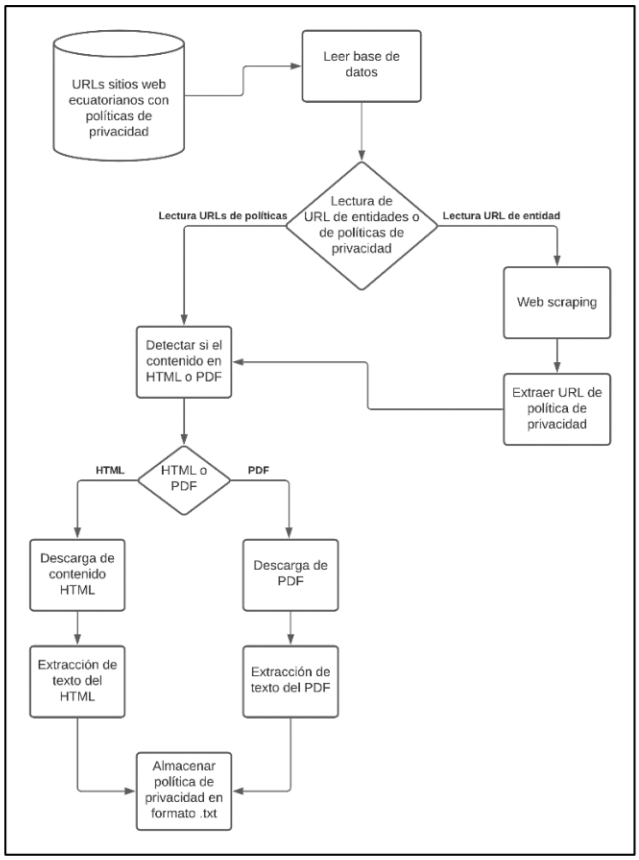


Figura 2.5. Diagrama de extracción de políticas de privacidad

El proceso de extracción de políticas de privacidad se fundamenta en dos pilares; el primero, la descarga de las políticas de privacidad mediante el enlace almacenado en la base de datos correspondiente, específicamente en la columna "*Policy URL*"; el segundo es un proceso más complejo, que implementa *web scraping* del enlace de la entidad (*Entity URL Full*) hasta encontrar el enlace a la política de privacidad y luego proceder a su descarga.

El proceso de extracción de las políticas de privacidad mediante *web scraping*, hace uso de la librería *Selenium* [29], la cual es un explorador web (*webdriver*) que permite configurar todos los parámetros de navegación mediante programación. Para el presente trabajo, se utiliza *geckodriver* [30], una API que permite realizar las consultas HTTP. Esta API es el núcleo de exploradores web tradicionales como Firefox.

Con el propósito de automatizar completamente el proceso de extracción, se programó al *web driver* para que se ejecute sin iniciar interfaz gráfica, lo cual permite agilitar el proceso y automatizarlo con mayor facilidad. Adicionalmente, debido a que el proceso de búsqueda de enlaces a políticas de privacidad se realiza iterativamente para todas las entidades en este escenario, se programó un tiempo de espera entre búsquedas de 2.4 segundos con el fin de que el proceso no sea detectado como un *bot.*

El proceso de recolección de políticas de privacidad se fundamenta en que el *webdriver* haga una petición al sitio web de la entidad mediante el enlace registrado y que, dentro del contenido HTML que obtenga, encuentre los hipervínculos existentes. Luego, entre los hipervínculos encontrados, el *webdriver* detectará aquellos potencialmente relacionados con una política de privacidad. Una vez obtenidos los enlaces, el *webdriver* accederá a ellos "dando clic", y obtendrá el contenido de la política de privacidad almacenado.

Durante la automatización de este proceso se encontraron dos problemas: el primero se generaba cuando la política de privacidad se abría en una nueva pestaña del explorador web, lo cual se solucionó detectando nuevas pestañas abiertas y cerrando aquellas no relacionadas con la política de privacidad; así, la descarga de la política de privacidad era posible. El otro problema encontrado se producía cuando un banner flotante se desplegaba en la página de la entidad; esto ocurría en la mayor parte de páginas web estatales, donde aparece un banner referente a la pandemia generada por el COVID-19. Para solucionar este inconveniente se programó al *webdriver* para que detecte el botón de cerrar presente en los banners. Debido a que no se sigue ningún tipo de estandarización para la asignación de nombres a las variables, la detección del botón generaliza el cierre del banner para los casos analizados.

Una vez que se cuenta con los enlaces de las políticas de privacidad, el método de extracción es similar en los dos casos (con y sin *web scraping*). El proceso de extracción es iterativo para cada uno de los enlaces a las políticas de privacidad. Debido a que los enlaces de las políticas de privacidad pueden apuntar a documentos HTML o documentos PDF, inicialmente se detecta esto en el enlace para según esto, proceder con su extracción.

En cualquiera de los casos descritos en el párrafo anterior, la obtención de la política de privacidad, se utiliza la herramienta *Requests* [31], la cual permite generar solicitudes HTTP de manera más simple que *Selenium.* La versatilidad en la configuración de *Requests* es menor, pero suficiente para descargar el contenido de las páginas web que contienen las políticas de privacidad.

Las conexiones recurrentes podrían ser bloqueadas por servicios que previenen ataque DDoS, debido al acceso automatizado que se implementa. Para evitar ese bloqueo, se añade una cabecera a las solicitudes indicando que la consulta se realiza desde un explorador web tradicional, utilizado en un computador con sistema operativo Windows 10, además de permitir codificación.

Cuando la política que se busca descargar se encuentra en formato HTML, por lo ligero del formato, se establece un tiempo de espera para dicha descarga de 3.5 segundos tras realizar la conexión con el servidor. Además, para el posterior análisis de las políticas, se decodifica el contenido HTML a formato UTF-8, este último usado para decodificar el contenido a castellano. El contenido HTML que se extrae es el texto correspondiente a la política de privacidad, es decir, se ignora todas las imágenes, enlaces y tablas.

Al detectar que la política de privacidad se encuentra almacenada en el servidor web como un archivo PDF, el proceso de extracción del texto de la política de privacidad es un tanto más complejo. A continuación, se presenta el diagrama del proceso, implementado utilizando la librería *PDFMiner* [32].

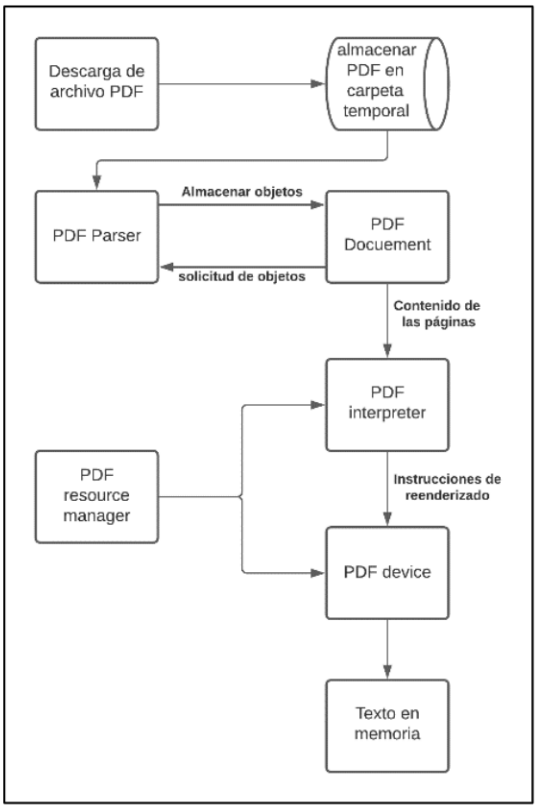


Figura 2.6. Diagrama de extracción de texto con PDFMiner [33]

En la Figura 2.6 se aprecia que el documento PDF con la política de privacidad es almacenado temporalmente en disco con la finalidad de utilizar minería de texto aplicada a documentos PDF. La minería de texto usada conlleva mayor dificultad en un documento PDF debido a que su estructura interna es mucho más compleja que la del texto claro. Los objetos que forman parte del proceso de extracción de texto, a partir de un documento PDF, se describen a continuación:

1. *PDF parser*: es el encargado de la comunicación con el archivo PDF. Obtiene los datos directamente del archivo como una secuencia de bits.

2. *PDF document*: almacena en memoria la estructura de los datos obtenidos del documento, para facilitar su lectura e interpretación.

3. *PDF interpreter*: analiza el contenido de cada una de las páginas del archivo PDF.

4. *PDF resourse manager*: almacena los recursos del archivo PDF como: imágenes, color del texto, tipo de letra, formularios, etc.

5. *PDF device:* transforma el contenido analizado de cada página en aquello que se necesita para el proceso. En este caso el contenido analizado se convierte a texto claro para su almacenamiento.

Finalmente, a partir de los dos tipos de archivos, se procede a almacenar cada política de privacidad como texto claro en formato .txt. El nombre utilizado para almacenar cada política de privacidad es el correspondiente al extraído de la columna *Entity URL* de la base de datos generada en la sección anterior.

2.3 GENERACIÓN DE BASE DE DATOS DE POLÍTICAS DE PRIVACIDAD

Como se indicó en la sección anterior, en este punto del trabajo realizado se cuenta con políticas de privacidad almacenadas en texto claro dentro de archivos con formato .txt.

Es común que los documentos de las políticas de privacidad incluyan "ruido" (*banners* de los sitios webs, menús, descripciones de imágenes, etc.). Esta información no relacionada con la política de privacidad podría provocar que los resultados obtenidos de su análisis sean inexactos. Así, es necesario realizar la "limpieza" de los documentos de las políticas de privacidad, actividad que permitirá, entre otras cosas, que el conteo de palabras y oraciones, así como la medición de su prevalencia se relacione únicamente con la política de privacidad.

Existen varias entidades que, debido al mal diseño de su sitio web, o por la naturaleza propia de este sitio web, presentan mucho ruido junto con la política de privacidad. Un ejemplo extremo de esto es la política de privacidad de Foros Ecuador [34] que es una política que está en línea con la normativa legal vigente en el momento de su publicación (septiembre de 2015), pero que, debido a la naturaleza del sitio web (un blog), esta política se creó como una entrada de blog que permite respuestas de usuarios tal como se aprecia en la Figura 2.7. Estas respuestas (muchas veces alejadas completamente del tema) distorsionarían el análisis que se haga de la política de privacidad.

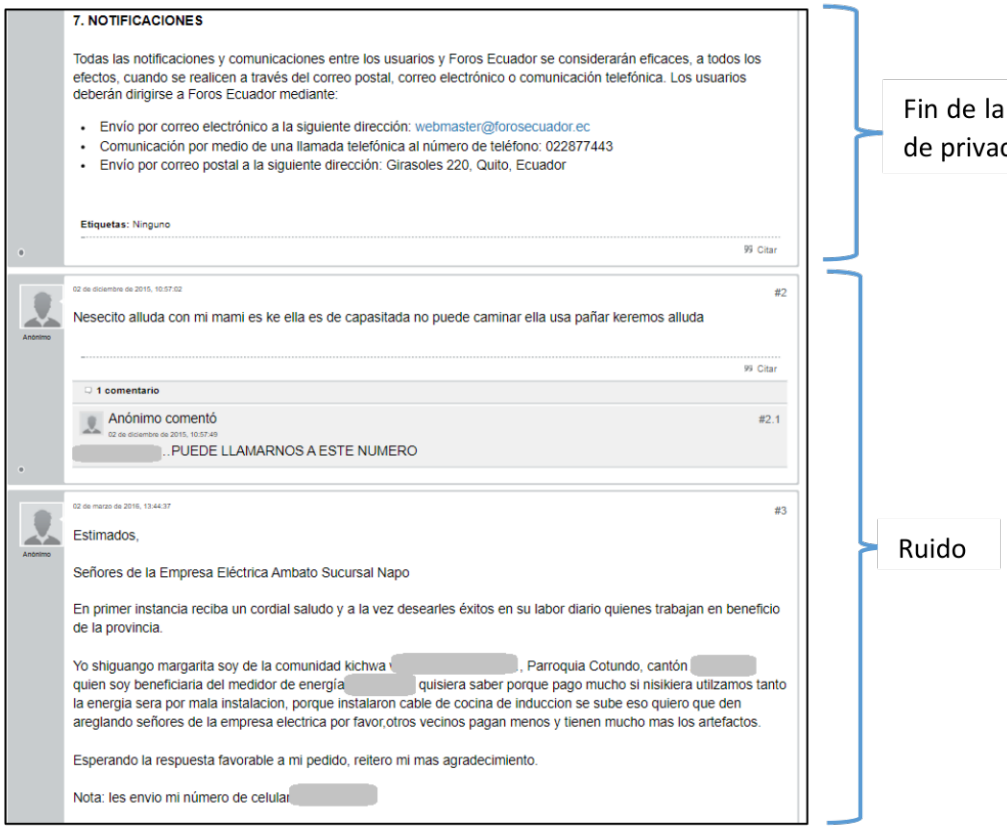


Figura 2.7. Ruido en política de privacidad de Foros Ecuador

Para realizar la limpieza de las políticas de privacidad, se realizó un *script* de Python que detecta en el documento un título que podría estar relacionado con una política de privacidad. A partir de la ubicación de ese título, se extrae el texto posterior, que se asume como el texto final correspondiente a la política de privacidad, y se procede a almacenar en un nuevo documento con formato .txt para su posterior análisis.

Debido a que se intenta medir la efectividad de este filtro, también se realiza una limpieza manual, para compararla con el mecanismo automatizado. Esto implica eliminar el ruido de la política de privacidad presente, tanto al inicio como al final de la política de privacidad.

Una vez que se han aplicado los filtros automatizado y manual a las políticas de privacidad, se procede a evaluar la eficiencia del filtro automatizado comparando la cantidad de palabras de la política obtenidas con ambos filtros. La razón de estas cantidades permite calcular el porcentaje de limpieza con la fórmula

$$Eficiencia = \frac{\#palabras\_filtro\_manual}{\#palabras\_filtro\_automatizado} * 100. \quad (2.1)$$

Tras evaluar la eficiencia del filtro, se procedió a almacenar las políticas de privacidad en diferentes directorios para facilitar su posterior análisis. Para esto, fue necesario el uso de la base de datos generada en la sección 2.1, que cuenta con la columna *Category,*

correspondiente la categoría en la cual fue clasificado cada sitio web de donde se extrajo la política. Para almacenar las políticas de acuerdo con esta categoría, tras extraer la información de los campos *Category* y *Entity URL* para cada política, se analiza el nombre con el cual fue almacenada dicha política. Debido a que el nombre usado para almacenar cada política es el mismo que el *Entity URL* es posible parear políticas con categorías y así moverlas al directorio designado para estas.

También se almacenan estas políticas en un repositorio público en GitHub [35] permitiendo así contar con una base de datos en línea, que podría inspirar futuros trabajos, por ejemplo, para analizar la evolución de las políticas de privacidad tras pasar el periodo de espera para la aplicación de la LOPDP.

GitHub fue seleccionada como la plataforma para almacenar las presentes políticas de privacidad para que cualquier usuario que desee copiar el repositorio pueda hacerlo utilizando la herramienta de consola `git` o descargando directamente desde la página web. Adicionalmente, en caso de requerirse, se podrá asignar permisos a quién lo requiera para escribir en el repositorio, pudiendo añadir las nuevas políticas de privacidad en función de sus actualizaciones.

Tras crear una base de datos con políticas de privacidad en texto claro y en formato .txt y proceder a realizar con su análisis, es necesario contar con una base de datos mucho que se pueda analizar más eficientemente. Para ello, se genera una estructura de datos utilizando la librería `pandas`.

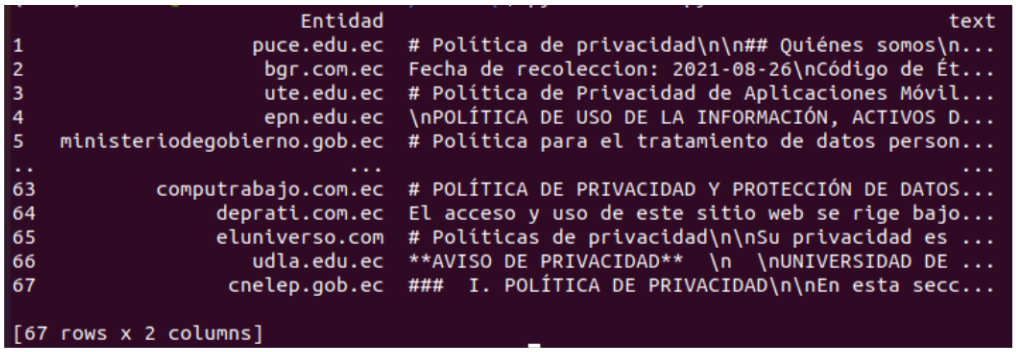
En la actualidad, `pandas` es la librería más utilizada para ciencia de datos en Python; al ser una librería de código abierto cuenta con un sin número de colaboradores que publican mejoras o documentación [36]. Esta documentación resultó de gran utilidad en el presente trabajo. La principal función de `pandas` dentro del presente trabajo se relaciona con la manipulación de tablas y documentos en formato .csv debido a las facilidades que ofrece, además de no requerir demasiado poder de procesamiento.

Mediante la biblioteca `pandas`, se procede a generar una estructura de datos llamada *DataFrame,* la cual es una estructura bidimensional de tamaño variable que tiene todas sus columnas y filas etiquetadas [37]. Para el presente trabajo se creó un *DataFrame* de 67 filas y 2 columnas, las columnas corresponden a la siguiente información:

- Entidad: contiene el nombre de la entidad (empresa, institución, etc.), que sirve para identificar a cada política de privacidad.
- Texto: contiene todo el texto de cada política de privacidad.

Para cada una de las políticas de privacidad analizadas se genera una entrada o fila en este *DataFrame*, permitiendo así tener una estructura de datos completa de las políticas para facilitar su análisis. La estructura de datos seleccionada ofrece facilidad en su manipulación y operación, además de que permite una extracción sencilla de información, que es clave para los siguientes pasos de nuestro análisis. En resumen, este análisis puede ser más eficiente y requerir así menos recursos del computador al ejecutarse.

En la Figura 2.8 se presenta un resumen del *DataFrame* generado para las políticas de privacidad. Se observan las 67 entradas, que incluyen los nombres de las entidades y los textos iniciales de cada política de privacidad.



	Entidad	text
1	puce.edu.ec	# Política de privacidad\n\n## Quiénes somos\n...
2	bgr.com.ec	Fecha de recoleccion: 2021-08-26\nCódigo de Ét...
3	ute.edu.ec	# Política de Privacidad de Aplicaciones Móvil...
4	epn.edu.ec	\nPOLÍTICA DE USO DE LA INFORMACIÓN, ACTIVOS D...
5	ministeriodegobierno.gob.ec	# Política para el tratamiento de datos person...
..	...	...
63	computrabajo.com.ec	# POLÍTICA DE PRIVACIDAD Y PROTECCIÓN DE DATOS...
64	deprati.com.ec	El acceso y uso de este sitio web se rige bajo...
65	eluniverso.com	# Políticas de privacidad\n\nSu privacidad es ...
66	udla.edu.ec	**AVISO DE PRIVACIDAD** \n \nUNIVERSIDAD DE ...
67	cnelep.gob.ec	### I. POLÍTICA DE PRIVACIDAD\n\nEn esta secc...

[67 rows x 2 columns]

Figura 2.8. *Dataframe* generado con políticas de privacidad

Finalmente, el *DataFrame* fue almacenado como un documento .csv con el propósito de cargarlo cada vez que sea necesario en las secciones descritas a continuación. Debido a que se utiliza la función `to_csv` [38] propia de la librería de `pandas`, el archivo almacenado contiene toda la información necesaria para que, al cargar el *DataFrame* nuevamente con la función `read_csv` [39], su estructura se mantenga, permitiendo así mantener todos los beneficios de utilizar un *DataFrame*.

2.4 ANÁLISIS CUANTITATIVO DE POLÍTICAS DE PRIVACIDAD

Tal como se describe en la sección anterior, las políticas de privacidad se encuentran en dos bases de datos diferentes (como archivos en texto claro y *DataFrame*) por lo cual en esta sección se utilizarán estas bases de datos para proceder con el análisis cuantitativo. Este análisis cuantitativo consiste en contar la cantidad de palabras y oraciones presentes en cada política de privacidad, contar las políticas de privacidad que cuentan con más de 3000 palabras, y finalmente medir la similitud entre las políticas de privacidad.

Los *scripts* desarrollados y descritos en esta sección se basan en la librería `NLTK`, que es una plataforma líder en análisis de lenguaje natural con Python. Como se apreció en la sección 1.4.4.2, el procesamiento de lenguaje natural cuenta con varios subprocesos como

clasificación, tokenización, creación de “corpus” (diccionarios de tokens), detección de tópicos, entre otros. Todos los subprocesos descritos pueden ser realizados utilizando la librería `NLTK` ya que esta plataforma recopila las principales librerías de PLN y las unifica para facilidad del desarrollador [40].

El análisis cuantitativo de las políticas de privacidad inicia con la obtención de la cantidad de palabras en cada política de privacidad con la finalidad de apreciar la longitud de la política de privacidad. Esta métrica podría servir, por ejemplo, para estimar la dificultad de la lectura de una política de privacidad muy extensa que, a su vez, incidiría en la comprensión del usuario.

Mediante la librería `NLTK`, utilizando el módulo `corpus.reader`, se generar un diccionario de palabras tokenizadas mediante la detección de espacios en blanco dentro de un texto plano. El módulo obtiene como entrada cada una de las políticas en texto claro, a partir de los archivos .txt, para luego procesarlas y almacenar en memoria el diccionario generado. Una vez que dicho diccionario se encuentra en la memoria, es posible extraer información relevante del diccionario. En el presente trabajo se extrae el número de palabras presentes en cada política de privacidad y adicionalmente la cantidad de oraciones presentes en cada política.

Para obtener el número de oraciones presentes en cada política; el módulo utilizado, detecta el inicio de una oración al encontrar una palabra que inicie con una letra mayúscula. Luego detecta la siguiente palabra más cercana que finalice en un punto. Al identificar esta estructura, el módulo considera que es una oración y la agrega al contador.

Este conteo de palabras y oraciones se realizó tanto para las tres versiones de cada política de privacidad: para la política descargada directamente sin filtrar, para la política filtrada manualmente, y para la filtrada de manera automatizada. Esto nos permite apreciar la magnitud en la variación del contenido en cada política, así como la eficiencia del filtro.

Con el fin de exponer la variación en la cantidad de palabras presentes en las políticas de privacidad analizadas, se representarán estas variaciones utilizando diagramas de barras.

Posteriormente, tras realizar el conteo de palabras y oraciones se procedió a verificar la similitud entre políticas de privacidad. Aprovechando las herramientas disponibles para esto, la detección de similitud podría ayudar a identificar algún patrón de comportamiento en la elaboración de este documento entre las entidades analizadas. Para realizar este análisis se siguió el procedimiento descrito en Figura 2.9.

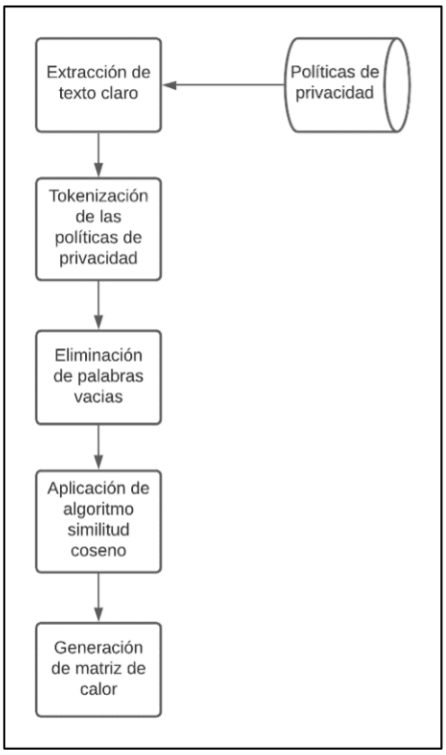


Figura 2.9. Proceso para cálculo de similitud

Tal como se describe en la figura anterior, las políticas de privacidad en texto claro almacenadas en el disco forman la base de datos usada como entrada para el presente proceso. las bases de datos usadas son dos, la base de datos con todas las políticas de privacidad y la base de datos con políticas de privacidad almacenadas en función de su categoría.

El proceso, al igual que el descrito para obtener la cantidad de palabras y oraciones, se fundamenta en la librería `NLTK`. Inicialmente, y tal como se describe en la sección 1.4.4.2, el texto de las políticas de privacidad es tokenizado para generar un corpus de cada política, facilitando así el uso de la librería. A continuación, mediante los diccionarios presentes en `NLTK`, se exporta el diccionario de palabras vacías (*stop words*) del idioma español y se las elimina de la política de privacidad.

Las palabras vacías son aquellas palabras que por sí solas carecen de significado, por lo que no poseen información relevante para el PLN; en español corresponden a los artículos, conjunciones, pronombres, preposiciones, etc. Aunque carecen de significado en el procesamiento automatizado, son necesarias para dar contexto o estructura al texto natural [23].

Luego de eliminar las palabras vacías del “corpus” (diccionario de palabras *tokenizadas*), se cuenta con todo lo necesario para calcular la similitud entre dos políticas de privacidad:

para ello, se usa la fórmula de la similitud coseno, que permite calcular una métrica de la similitud que existe entre dos vectores.

En el caso del presente trabajo los vectores se encuentran formados por palabras tokenizadas (representaciones numéricas de cada palabra). Por ejemplo, los vectores $A = [hola, mundo]$ y $B = [saludos, mundo]$, tras ser tokenizados son representados como $A = [1, 2]$ y $B = [3, 2]$.

El fundamento matemático radica en el hecho de que

$$\cos 0^{\circ} = 1 \tag{2.2}$$

y que

$$\cos x = [0{,}0.99] \tag{2.3}$$

$$1^{\circ} <= x <= 90^{\circ}. \tag{2.4}$$

La similitud entre dos vectores se puede interpretar en función del ángulo entre ellos (o su coseno, tal como se indica en las ecuaciones 2.2 y 2.3). Así, si el coseno del ángulo es 1 (0º entre los vectores), se tratará del mismo vector (100% similares), y mientras más cercano a 0, más similares. Igualmente, si los vectores no son similares, el coseno del ángulo entre ellos será menor que 1, por lo que su similitud estará entre el 0% y el 99% [41].

Para calcular el coseno del ángulo entre dos vectores ($A\ y\ B$), se recurre a la fórmula del producto escalar entre ellos, que se define como el producto de sus módulos por el coseno del ángulo comprendido entre estos. A continuación, se detalla la fórmula del producto escalar.

$$A \cdot B = \|A\| * \|B\| * cos(\theta) \tag{2.4}$$

Despejando la ecuación 2.4 para obtener $\cos\theta$:

$$cos(\theta) = \frac{A \cdot B}{\|A\| * \|B\|} \tag{2.5}$$

Tras reemplazar en la ecuación 2.5 la fórmula del módulo de un vector y la fórmula de producto escalar, se obtiene la siguiente fórmula:

$$cos(\theta) = \frac{\sum_{i=0}^{n} A_i B_i}{\sqrt{\sum_{i=0}^{n} A_i^2} * \sqrt{\sum_{i=0}^{n} B_i^2}} \tag{2.6}$$

Finalmente, la ecuación 2.6 es la que se utiliza para calcular la similitud [42] entre las diferentes políticas de privacidad. Ya que el resultado de esta fórmula se encuentra entre 0 y 1, se multiplica por 100 para hallar un porcentaje que facilite su comprensión.

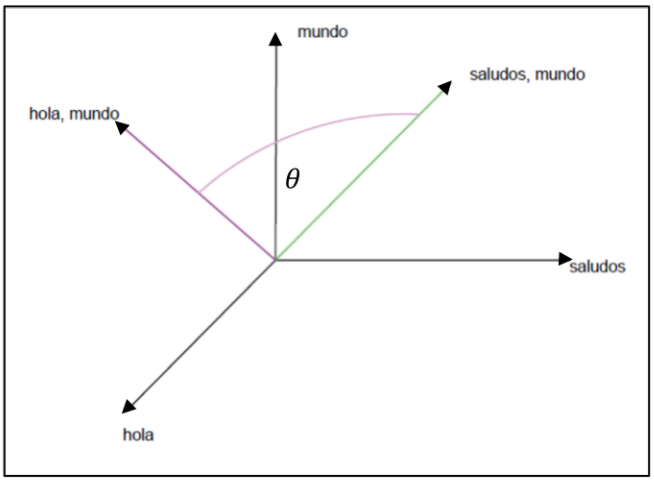


Figura 2.10. Proyección tridimensional de similitud coseno

En la Figura 2.10 se aprecia una representación tridimensional de cómo se calcula la similitud coseno en textos; en este caso los ejes son las palabras: saludos, hola y mundo; posteriormente la oración “hola, mundo” y “saludos, mundo” son graficadas como vectores en los planos correspondientes. También se identifica el ángulo θ entre esos vectores. Esta representación permite comprender la distribución que tendría el texto de las políticas de privacidad como vectores para poder aplicar la ecuación 2.6.

Las similitudes entre los documentos se organizaron en un *DataFrame* de `pandas` para facilitar su procesamiento. El resultado final es una serie de matrices (para diferentes categorías de sitios web) que registran las similitudes de las políticas de privacidad entre cada par de entidades estudiadas.

Debido a que 67 políticas de privacidad resultan complicadas de visualizar en un gráfico, se muestran en estas matrices de similitud solamente las más relevantes; en este caso aquellos pares de políticas que tienen más de 30% de similitud. Adicionalmente, se genera un *DataFrame* para cada una de las categorías, permitiendo apreciar similitud dentro de la misma categoría. Estos *DataFrame*s son almacenados en formato .csv para facilitar su uso posterior ya que el cálculo de similitud es un proceso que requiere gran poder computacional por lo que su ejecución se busca realizar una única vez.

A partir de la generación de las matrices de similitud, se representa dicha solicitud mediante mapas de calor que facilitan la interpretación de los resultados. Para ello, se usa la librería

`seaborn` [43], que permite generar gráficos avanzados para visualización de datos estadísticos, y particularmente la función `heatmap()`.

2.5 ANÁLISIS DE LA FRECUENCIA DE PALABRAS EN POLÍTICAS DE PRIVACIDAD

La cantidad de palabras presentes en cada política de privacidad, como se explica en la sección 2.4, podría dar una pista de la dificultad de leer estos documentos.

Por otra parte, se analiza también la frecuencia de palabras utilizadas en estos documentos, como un mecanismo para caracterizar su contenido. Esto ayudaría, por ejemplo, a identificar ciertas temáticas abordadas que no necesariamente se relacionen con el objetivo de una política de privacidad.

La frecuencia con la cual aparecen las diferentes palabras dentro de texto son un indicador del contenido del texto. Por ejemplo, si en un texto las palabras inteligencia y artificial aparecen con mayor frecuencia que otras, es posible inferir que el texto se refiera al aprendizaje de máquina. De esta manera, al analizar la prevalencia de las palabras en las políticas de privacidad, se podrían detectar palabras anómalas en el contexto de la protección de datos personales, y, por tanto, algo que se podría mejorar.

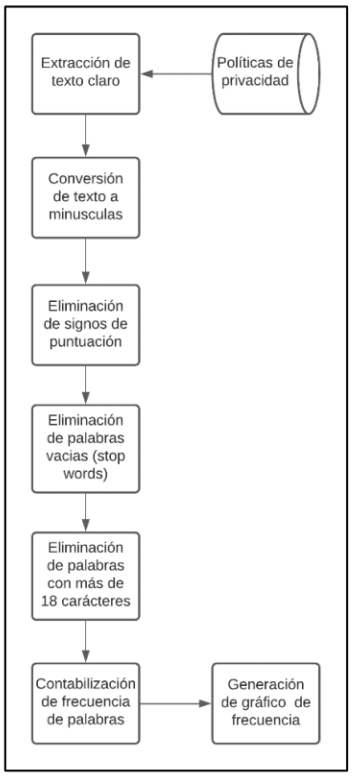


Figura 2.11. Proceso para análisis de frecuencia de políticas de privacidad

En la Figura 2.11 se ilustra el proceso para representar las palabras más frecuentes en todas las políticas de privacidad analizadas. La representación resultante es una nube de palabras en la que el tamaño de estas está en función de la frecuencia de aparición en las políticas. A continuación, se describe este proceso.

Inicialmente, se cargan todas las políticas de privacidad como texto claro desde el disco. Estas políticas son concatenadas en una sola cadena de texto para tener una idea general de la prevalencia en todas las políticas analizadas.

Posteriormente, se convierten todas las palabras a letras minúsculas para facilitar el procesamiento del texto. Así, una misma palabra escrita en mayúsculas o en minúsculas contará como una sola palabra.

Una vez que se cuenta con la cadena de texto que contiene todas las políticas de privacidad escritas con letras minúsculas, se eliminan todos los signos de puntuación. Para ello se crea un diccionario con estos signos de puntuación, que sirve de referencia para buscarlos y eliminarlos de la cadena de texto.

Tal como se explicó en la sección 2.4, es necesario eliminar las *stop words* del texto, debido a que no aportan información al análisis realizado. Por este motivo se recurre a la librería `NLTK` que contiene el módulo del diccionario de palabras vacías; en este caso se usa el diccionario correspondiente a español.

Adicionalmente, también para facilitar el procesamiento del texto, se programa un filtro para eliminar palabras con más de 18 caracteres. Estas palabras son poco comunes y no se espera que aparezcan en el contexto de este análisis. Además, palabras tan extensas podrían ser producto de errores ortográficos. Este número de caracteres máximo se escogió en función de un análisis de las palabras más extensas en el idioma español [44].

Una vez se ha preprocesado (“limpiado”) el texto de las políticas, se determina la frecuencia de cada palabra. Ya que la cadena de texto tiene un gran tamaño, y con propósitos de optimización de poder computacional, se recurre al uso de *hash tables* permitiendo que el script desarrollado sea eficiente. Tras contabilizar la frecuencia de aparición de cada palabra, se obtiene una *hash table* en la que cada “llave" corresponde a una palabra y su “valor” a la cantidad de veces que esta aparece.

Finalmente, con la *hash table* creada, es posible realizar un gráfico que permita comprender de manera visual la frecuencia de aparecimiento de cada palabra. Para esto se recurre a la librería `WordCloud` [45], que grafica las palabras que aparecen con mayor frecuencia; adicionalmente, le asigna un mayor tamaño a aquellas que aparecen más frecuentemente.

Para generar un mayor impacto en el gráfico desarrollado, se seleccionó el gráfico de una nube como máscara sobre la cual se graficaron las palabras. La nube ha sido seleccionada, debido a que las políticas de privacidad permiten conocer, entre otras cosas, el procesamiento que se realiza de los datos personales recopilados, y este procesamiento suele realizarse en su gran mayoría en servicios en la nube.

2.6 MODELADO DE TEMAS EN POLÍTICAS DE PRIVACIDAD

Tal como se explicó en la sección 1.4.4.2, el modelado de temas es una de las principales aplicaciones del PLN. Se trata de una técnica de aprendizaje de máquina no supervisada orientada a la clasificación. Como parte de este trabajo, se aplica el modelado de temas con el fin de agrupar las diferentes políticas de privacidad en diferentes temas o tópicos y obtener otra caracterización de estos documentos. Podría esperarse, por ejemplo, que, entre los tópicos encontrados, se manifiesten los distintos elementos que deberían formar una política de privacidad (declaración de derechos del usuario, tipo de procesamiento que se realizará, terceras partes involucradas, etc.).

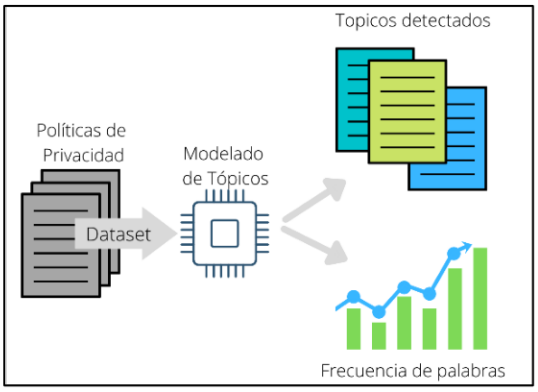


Figura 2.12. Modelado de tópicos [46]

Este modelado de tópicos para las políticas de privacidad se realiza determinando el número de tópicos óptimos para los documentos analizados. Estos tópicos estarán representados por las palabras más relevantes asociadas a cada uno. Esto permitiría verificar si las políticas de privacidad analizadas tratan temas relativos a lo estipulado por la LOPDP en su artículo 12 (sección 1.4.2.2).

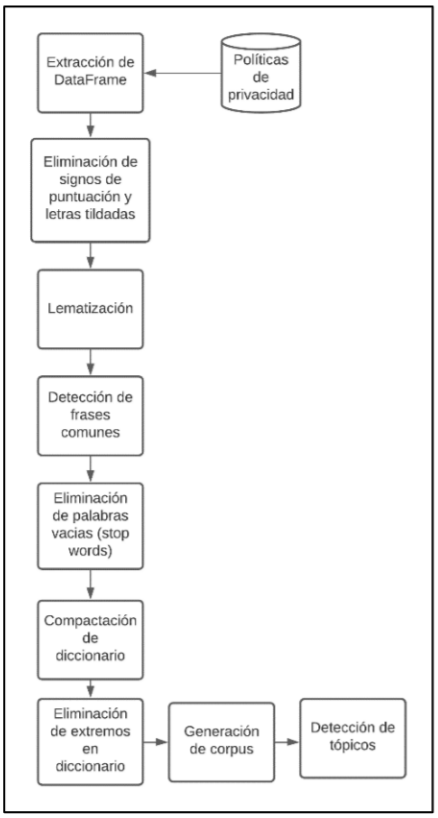


Figura 2.13. Modelado de tópicos en políticas de privacidad

Tal como se describe en la sección 2.3, el uso de un *DataFrame* permite agilitar el proceso de análisis de lenguaje natural por lo que para la presente sección se hará uso de esta base de datos.

En la Figura 2.13 se observa el proceso realizado para el modelado de tópicos en políticas de privacidad. Una vez cargada la base de datos, se procede a eliminar los signos de puntuación de las políticas de privacidad, y a reemplazar todas las palabras que se encuentren tildadas; esto con la finalidad de simplificar el trabajo de las librerías utilizadas pues están diseñadas para procesar texto en idioma Inglés.

Una vez que se cuenta con la base de datos filtrada, se procede a utilizar la librería `gensim` [47], diseñada para Python y orientada al procesamiento de lenguaje natural; sus funciones incluyen el modelado de tópicos, la indexación de documentos y el análisis de similitud para corpus de gran tamaño.

Con la finalidad de procesar únicamente las palabras que sean realmente necesarias y no repetir términos, se recurre a la función `lemmatization()`, que se encarga de realizar el proceso de lematización. La lematización consiste en hallar el lema o raíz correspondiente para cada palabra a partir de una forma flexionada de la misma (es decir,

el femenino, conjugación, el plural, etc.); al hacerlo, se obtiene un número menor de palabras en el diccionario final y esto permite contar adecuadamente la frecuencia de aparición de las diferentes palabras. Para ilustrar este proceso, por ejemplo, las palabras *usuario, usuaria, usar, usará*, tendrían el mismo lema, *usar* [48].

Luego de lematizar el texto de las políticas de privacidad, se detectan las frases más comunes que son conjuntos de oraciones que, al estar estrechamente asociadas, se tratan como una sola. Con ese fin se utiliza la función `models.phrases()` [49]. Por ejemplo, la frase "política de privacidad", al estar formada por 3 palabras que aparecen juntas de manera frecuente, se consideran como una solo palabra; esto permite tener en cuenta estructuras más complejas dentro de la detección de tópicos haciendo que esta sea más cercana a la realidad.

Tal como se indicó en las secciones 2.4 y 2.5, la eliminación de *stop words* es un paso indispensable para el PLN. Así, esta se implementa previo a la detección de tópicos y para esto se eliminan las *stop words* presentes en el diccionario de palabras vacías de la librería `NLTK`.

Al eliminar las *stop words* del diccionario, se crean brechas en sus *hash tables*, que causan errores al momento de analizar tópicos. Para cubrir esas brechas, se usa la función `Dictionary.compactify()` de la librería `gensim` [50].

Una vez que se cuenta con un diccionario completamente limpio, se eliminan de este palabras raras o poco frecuentes; por ejemplo, alguna (como un nombre propio) que tiene sentido sólo el contexto de una institución. Este nuevo filtro se implementa con la función `filter_extremes()` [51], que se configura para eliminar las palabras que no se encuentren en al menos 2 documentos, y en no más del 97% de los documentos; esto permite, respectivamente, eliminar palabras con poca frecuencia que no aporten información, y eliminar palabras muy repetitivas que generen sesgo. En este punto también se utiliza la función `Dictionary.compactify()` para consolidar el diccionario luego de este proceso de filtrado.

Tras realizar el preprocesamiento descrito anteriormente, se procede a crear el diccionario que se utilizará como entrada para el algoritmo de modelado de tópicos en las políticas de privacidad. El diccionario que se generara es del tipo corpus (palabras tokenizadas), el cual tendrá duplas formadas por las palabras tokenizadas y la frecuencia con la que aparecen. Para generar este corpus se hace uso de la función `dictionary.doc2bow()` la cual permite agregar a un corpus final las duplas generadas para cada política de privacidad [52].

Una vez que se cuenta con los datos necesarios para modelar los tópicos de las políticas de privacidad de Ecuador, se procede a usar el algoritmo LDA mediante la función `ldaModel()` presente en la librería `gensim`. La selección de este modelo obedece a que cuenta con mayor documentación por parte de los desarrolladores de la librería y a que es el más popular en el campo del modelado de tópicos en textos extensos.

Tal cómo se describe en la sección 1.4.4.2, LDA permite el modelado de tópicos asignando pesos a estos tópicos y a las palabras para luego, mediante la distribución estadística de Dirichlet, asignar las palabras más relevantes a los tópicos. La cantidad de tópicos a modelar es una variable que debe definirse en base a una métrica ya que el seleccionar muchos o muy pocos tópicos podría sesgar el análisis.

Para la selección del número adecuado de tópicos, se utiliza un diagrama de coherencia contra el número de tópicos, tal como se aprecia en la Figura 2.14. Este diagrama permite apreciar de manera gráfica cual será el número óptimo de tópicos que, de la manera más coherente, permitan describir la información presente en las políticas de privacidad analizadas.

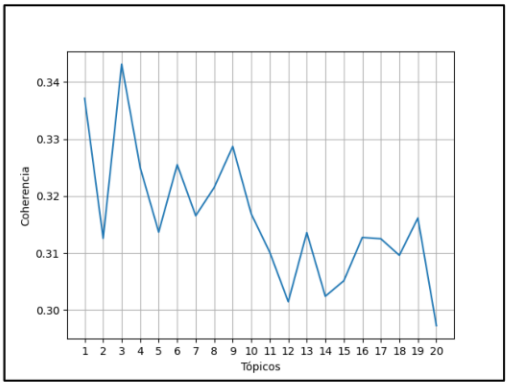


Figura 2.14. Coherencia vs. Número de Tópicos

En el contexto del modelado de tópicos, la coherencia mide qué tan relacionadas están las palabras pertenecientes a un tópico en cuanto a semántica se refiere [53]. Por ejemplo, las palabras: privacidad, política, tratamiento, datos, usuario están relacionadas coherentemente al campo semántico de una política de privacidad.

Finalmente, tras la selección del número adecuado de tópicos se procede a imprimir en pantalla los tópicos (conjunto de palabras más relevantes), junto con la coherencia alcanzada para dicho número de tópicos. Adicionalmente, se genera un archivo .html interactivo que permite apreciar las palabras más relevantes de cada tópico junto con el peso asignado para cada una de estas; permitiendo observar cómo se encuentran

distribuidos los tópicos en un plano para verificar si existen o no, tópicos que interfieren con el campo semántico de otros, esto se realiza mediante la librería `pyLDAvis()` [54].

2.7 CONTRASTE REALIZADO CON MÉXICO

Una vez que se analiza de manera automatizada las políticas de privacidad de los principales sitios web de Ecuador (según el ranking Alexa), se contrasta parte de los resultados obtenidos con aquellos obtenidos a partir de políticas de privacidad de otro país, México, que cuenta con una regulación de privacidad y protección de datos mucho más madura que la ecuatoriana.

Para poder realizar la comparación, se procedió a desarrollar el mismo procedimiento que para las políticas de privacidad ecuatorianas. Esto incluye las siguientes actividades.

1. Generación de base de datos de enlaces de políticas de privacidad para los principales sitios web de México según el ranking Alexa. El proceso realizado es similar al detallado en la sección 2.1. Al finalizar esta fase, se obtiene un documento en formato .csv con los enlaces a las políticas de privacidad, entidades y Wayback Machine, así como la entidad y categoría asociada a cada política.

2. Extracción de políticas de privacidad, en esta fase se realizó la descarga de las políticas de privacidad presentes en el documento .csv generado en la fase anterior; para llevar a cabo la presente fase, se realiza un proceso similar al descrito en la sección 2.2, donde al finalizar se obtienen los documentos de las políticas de privacidad en formato .txt.

3. Generación de base de datos de políticas de privacidad. Mediante las políticas de privacidad almacenadas en la sección anterior, se genera una base de datos mediante la categorización de políticas de privacidad en diferentes directorios; además, se genera la estructura de datos *DataFrame* para la facilidad al momento de procesar esta información. La tarea realizada es similar a la presente en la sección 2.3.

4. Al igual que en la sección 2.4, tras obtener las bases de datos, se procede a realizar el análisis cuantitativo de políticas de privacidad que se fundamenta en el conteo de palabras y oraciones, y el cálculo de similitud entre políticas de privacidad.

5. El posterior procesamiento realizado es el análisis de frecuencia de palabras en políticas de privacidad, tal como se explica en la sección 2.5. Esto permite caracterizar las políticas de privacidad e identificar el "ruido" existente en ellas.

6. Finalmente, se procede a realizar el modelado de temas presentes en las políticas de privacidad, para lo cual se realiza un proceso similar al descrito en la sección 2.6. El objetivo de esta actividad es conocer los principales temas que abarcan las políticas de privacidad para así poder detectar irregularidades presentes en el texto.

En base a los resultados obtenidos se compara las particularidades encontradas. Adicionalmente, se generan diagramas de barras que permitan observar las diferencias presentes entre las políticas de privacidad ecuatorianas y mexicanas.

2.8 ANÁLISIS MANUAL DE POLÍTICAS

Con el fin de contrastar el análisis automatizado de políticas de privacidad en Ecuador, se planifica un análisis manual de estos documentos. Este análisis parte de la colaboración de un grupo de voluntarios (estudiantes de la EPN, con ciertos conocimientos de privacidad) que llenan una encuesta luego de leer las políticas de privacidad procesadas.

Para que los colaboradores puedan completar la encuesta realizada se les proporcionó las 67 políticas de privacidad analizadas en el presente trabajo. Las preguntas realizadas buscan medir el cumplimiento de la LOPDP, y hasta cierto punto la dificultad de leer las políticas de privacidad. Las preguntas realizadas fueron:

- ¿Cuánto tiempo le tomó leer la política? (en minutos). El objetivo de esta pregunta es poder contrastar el tiempo promedio calculado mediante el análisis automatizado realizado y el tiempo real que tardar los usuarios en leer las políticas. Permitiendo analizar cómo este tiempo afecta en el objetivo primordial de una política de privacidad, el cual es informar al usuario.

  Las opciones de respuesta provistas en esta pregunta son:

  - menos de 5 minutos
  - entre 5 y 10 minutos
  - entre 10 y 15 minutos
  - más de 15 minutos

- ¿La política le pareció muy larga?. Esta pregunta es subjetiva ya que depende de la percepción de cada uno de los encuestados. Sin embargo, podría ayudar a contrastar los datos obtenidos en el análisis automatizado para inferir si la extensión de las políticas de privacidad afecta a su comprensión, tal como se plantea en la sección 1.4.3. Un texto de gran extensión puede resultar difícil de comprender.

  Las opciones de respuesta provistas en esta pregunta son:

  - Si
  - No

- ¿Cuál es la base legal que se menciona en la política? El motivo de esta pregunta es conocer si las políticas de privacidad analizadas especifican el marco legal en la cual se fundamentan. Se espera que las políticas cuenten con una de las leyes descritas en la sección 1.4.2 Idealmente las políticas de privacidad deberían contar con la LOPDP como base legal, pese a que su aplicación no es obligatoria en la actualidad.

Adicionalmente, se realiza un análisis exploratorio de las políticas de privacidad en busca de particularidades para ser analizadas en el contexto del presente trabajo. Los hallazgos podrían servir para realizar futuros trabajos en la línea del estudio de la privacidad en Ecuador, así como para generar recomendaciones de buenas prácticas al redactar políticas de privacidad.

3 RESULTADOS, CONCLUSIONES Y RECOMENDACIONES

3.1 RESULTADOS

3.1.1 SITIOS WEB CON POLÍTICAS DE PRIVACIDAD

A partir del análisis manual de 139 sitios web populares en Ecuador, se determina la ausencia de políticas de privacidad en un importante número de ellos. El 44.6% no cuenta con una política de privacidad. Esto es, sin duda, preocupante, porque revela el gran desinterés de las instituciones con respecto al derecho de los usuarios a estar informados sobre lo que se hace con sus datos. Cabe recalcar que estos sitios web tampoco cuentan política de cookies.

3.1.2 SITIOS WEB CON PROTECCIÓN PARA LA EXTRACCIÓN DE INFORMACIÓN

Debido a la gran cantidad de sitios web que existen, la automatización del proceso de recolección y análisis de políticas de privacidad es clave; por ejemplo, durante un proceso de auditoría que requiera hacer la Autoridad de Protección de Datos Personales, con el objetivo de comprobar el acceso al derecho de los usuarios a estar informados; esta automatización permitirá al equipo auditor la recolección y análisis de un enorme volumen de datos presente en las políticas de privacidad de las entidades.

Sin embargo, se pudo comprobar que esta automatización se puede ver limitada por mecanismos de protección frente a ataques de denegación de servicio implementados por los sitios web. Particularmente, se encontró que de 77 sitios web que contaban con políticas de privacidad, el 13% no permitieron la descarga automatizada de sus políticas de privacidad, usando los mecanismos propuestos.

3.1.3 LIMPIEZA AUTOMATIZADA DE POLÍTICAS DE PRIVACIDAD

A partir del filtro desarrollado para limpiar el “ruido” de los documentos de políticas de privacidad, se encontró que la política del sitio web frecuento.com es una de las que más contenido tiene que no está relacionado a una política de privacidad. En la figura del ANEXO I, se puede apreciar que más del 60% de las palabras no corresponden a la política de privacidad de esta entidad, sino a otro contenido, por ejemplo: cabeceras, publicidad,

etc. Este “ruido” podría dificultar la comprensión del documento, que en sí mismo ya es complejo.

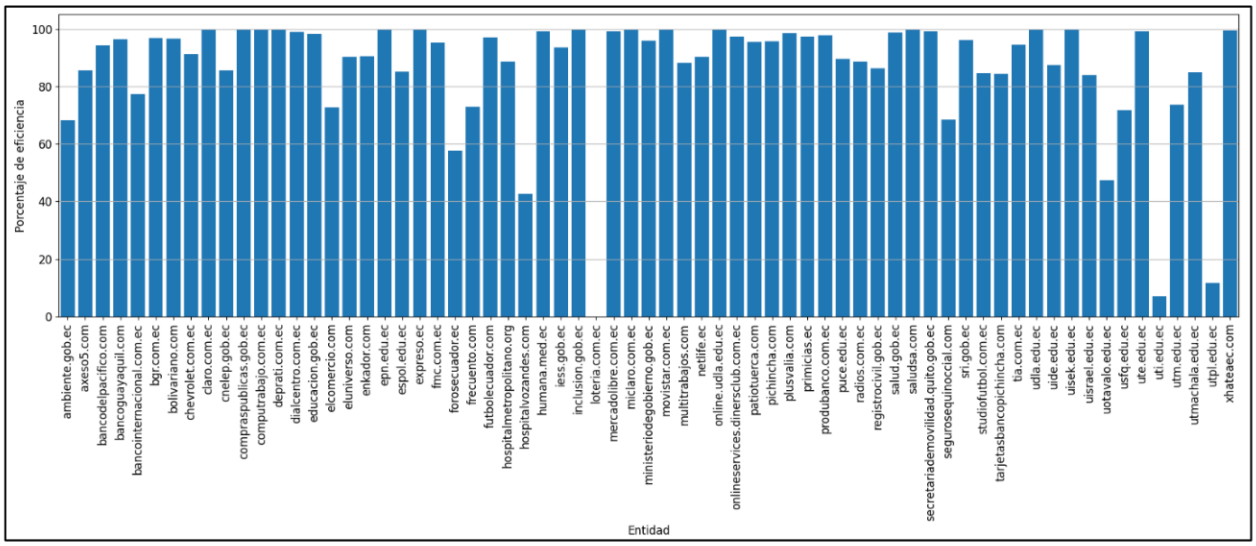


Figura 3.1. Eficiencia del Filtro aplicado para la limpieza automatizada

Al comparar la política de privacidad extraída manualmente y aquella limpiada de forma automatizada, se verifica la eficiencia del filtro, y esta se ilustra en la Figura 3.1, para cada sitio web. En el caso de loteria.com.ec, la eficiencia del filtro es 0% pues la política no tiene una marca al principio o al final del texto, que permita eliminar el ruido existente, como en los otros casos. Por otra parte, la eficiencia del filtro en frecuento.com es mayor al 70% luego de eliminar el 62% del texto. En general, en el 80% de políticas de privacidad, el filtro obtiene el 80% o más de eficiencia en la limpieza. Sólo en el 8.9% de las políticas de privacidad se obtiene una eficiencia de limpieza menor al 60%. En base a lo expuesto, se considera que la eficiencia del filtro es adecuada.

Haciendo un análisis general, en base a la tarea de limpieza, se encontró que en más del 10% de las políticas de privacidad analizadas, el 25% de esos textos correspondían a ruido, que podría hacer más difícil la lectura del documento a los usuarios.

3.1.4 CANTIDAD DE PALABRAS Y ORACIONES PRESENTES EN CADA POLÍTICA DE PRIVACIDAD

En la figura del ANEXO II se observa la cantidad de palabras por política de privacidad, etiquetadas según la entidad a la cual pertenecen. Se representa esta cantidad para 3 versiones de cada política: política sin limpiar, política limpiada manualmente, y política limpiada automáticamente.

La cantidad de palabras varía significativamente de entidad a entidad, y en los casos extremos esto podría dificultar que los usuarios ejerzan su derecho de acceso a la información. Por ejemplo, muy poco texto podría ser indicador de que no se informa detalladamente lo que manda el artículo 12 de la LOPDP. En contraste, si una política de privacidad tiene demasiado texto podría significar que contiene información no relevante para el usuario, o que contiene muchos términos y lenguaje legal que complicarían la lectura.

El promedio de palabras por política de privacidad es 2376 que, a un hispanohablante, le tomaría leer 10 minutos, si lee en promedio 250 palabras por minuto [55]. Además, aproximadamente el 24% de dichas políticas podrían ser leídas en 12 minutos (más de 3000 palabras). Estos tiempos son extensos, considerando que deberían dedicarlos a leer la política de privacidad de cada sitio que visitan. Esto sin duda desincentiva la lectura de este documento.

Al contar las oraciones (figura del ANEXO III), el promedio en cada política es 114, es decir, cada una cuenta con 21 palabras. Se trata de oraciones relativamente largas [56], más complicadas de leer. Este es otro problema de estos documentos cuyo objetivo es mantener al usuario informado.

Se encuentran políticas con pocas oraciones como la de “futbolecuador.com”, que es adecuada en relación con el marco legal previo a la LOPDP. Sin embargo, también existen políticas de privacidad con muchas oraciones como la de “mercadolibre.com.ec” que, siendo adecuada, podría resultar muy difícil de leer.

Finalmente, en las figuras de los ANEXO II y ANEXO III es posible observar la diferencia entre aplicar el filtro automatizado y el filtro manual para limpiar las políticas de privacidad. El filtro automatizado en general cumple su función correctamente, salvo en ciertos casos como la política de “uti.edu.ec”.

3.1.5 SIMILITUD ENTRE POLÍTICAS DE PRIVACIDAD

A continuación, se presentan mapas de calor que muestran la similitud entre las diferentes políticas de privacidad de las diferentes entidades. La similitud hallada permite comprender si las políticas son generadas siguiendo un formato común según la entidad o categoría específica.

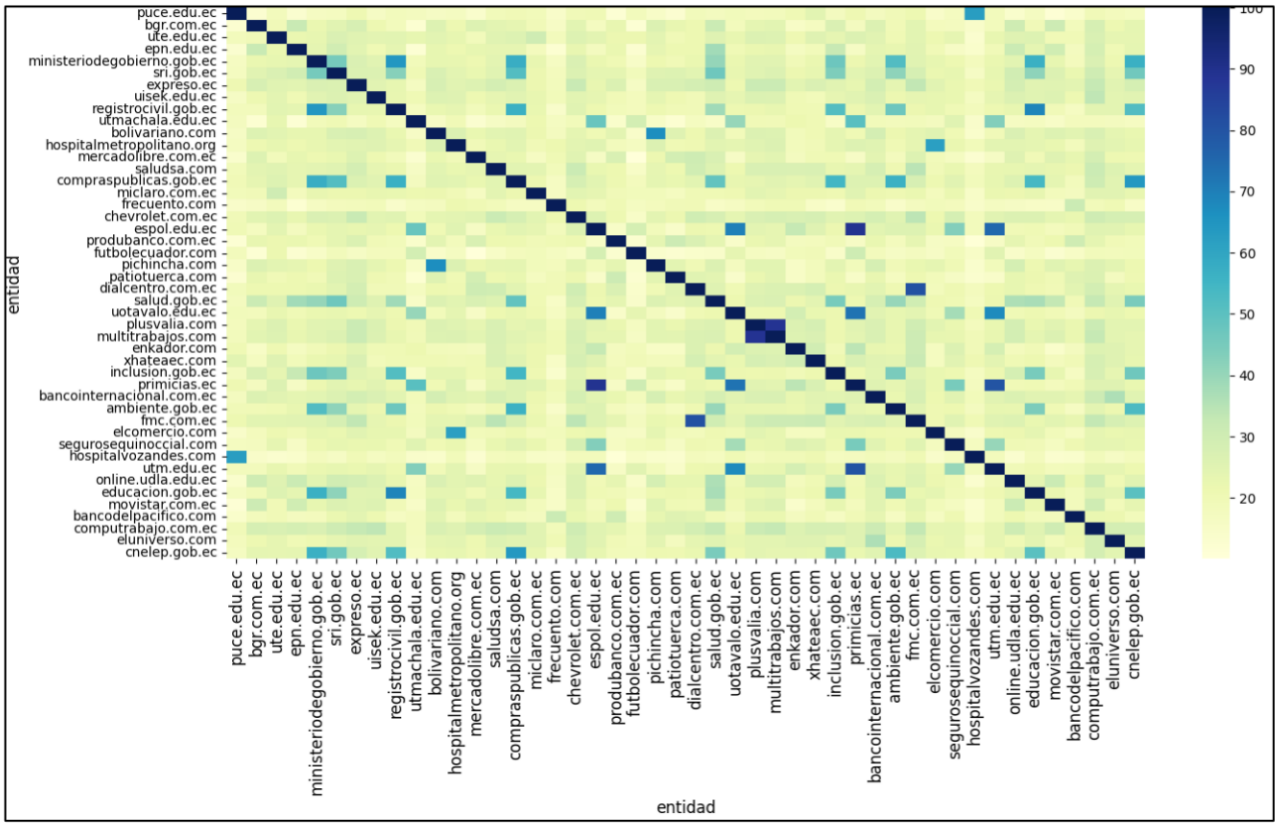


Figura 3.2. Similitud mayor al 30% en políticas de privacidad

Inicialmente, en la Figura 3.2 se aprecia las políticas de privacidad con al menos 30% de similitud con otra. De las 67 políticas de privacidad analizadas, 46 tienen más del 30% de similitud con alguna otra.

La similitud de políticas de sitios de la misma categoría es comprensible. Sin embargo, se encontraron ciertos casos en los cuales las políticas de privacidad de sitios de distintas categorías eran muy similares. Esto podría sugerir varias cosas, entre ellas que las entidades usen una misma plantilla para generar la política.

A continuación, se detallan estos casos.

- Las políticas de “primicias.ec”, que pertenece a la categoría *news (noticias),* las de “espol.edu.ec”, “uotavalo.edu.ec” y “utm.edu.ec”, son prácticamente las mismas (Figura 3.3). Esto puede deberse a que utilizan el mismo formato base o a que la institución encargada de generar estas políticas es la misma para las tres entidades.

  La similitud de las políticas en estos casos, al tratarse de entidades distintas, podría indicar que seguramente las declaraciones en ellas no reflejan la realidad de la recolección y procesamiento de los datos personales.

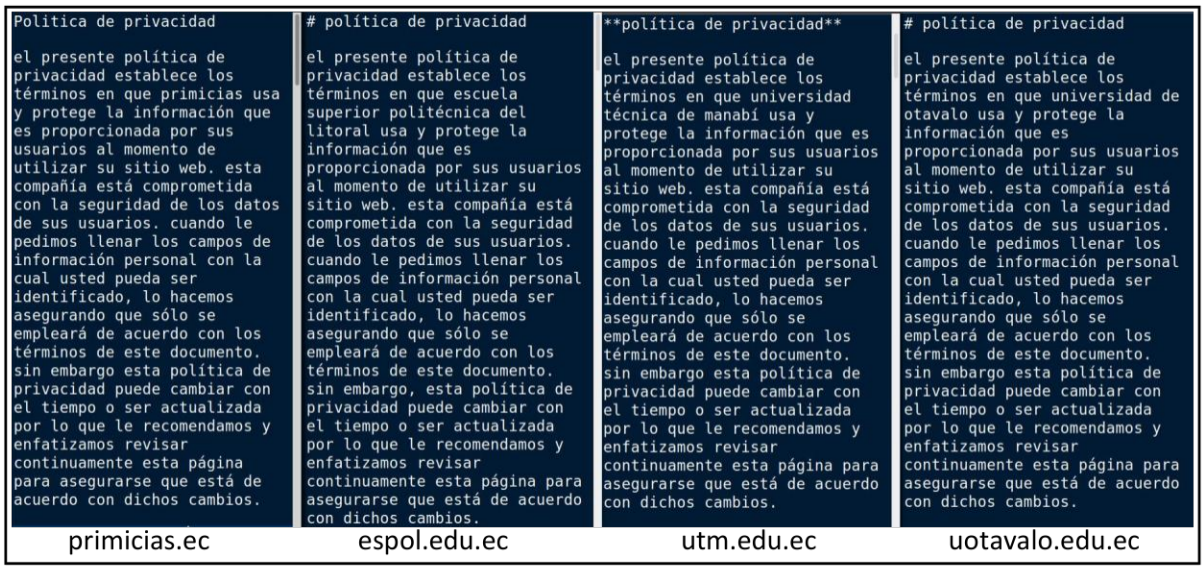


Figura 3.3. Similitud de políticas de privacidad de “primicias.ec”, “espol.edu.ec”, “uotavalo.edu.ec” y “utm.edu.ec”

- Asimismo, las entidades “multitrabajos.com” y “plusvalía.com” tienen políticas de privacidad 90% similares (ANEXO IV) aunque pertenecen a categorías distintas (empleo y compras, respectivamente).

- La entidad “puce.edu.ec”, que es una entidad perteneciente a la categoría *education (educación)* y la entidad “hospitalvozandes.com” perteneciente a la categoría *health* (salud), tienen un porcentaje de similitud mayor al 75% (ANEXO V). Pese a ser muy similares, el porcentaje de similitud no es mucho mayor debido a que con la entidad “hospitalvozandes.com” el filtro automatizado tiene una eficiencia del 40%.

También se analizaron los resultados de similitud entre políticas de privacidad de entidades de una misma categoría. En este caso podría ser comprensible que las políticas sean algo similares.

En la categoría *education* la mayor parte de políticas de privacidad tienen un porcentaje de similitud baja. Sin embargo, tal cómo se pudo apreciar en la Figura 3.3, existen varias entidades educativas que cuentan con políticas de privacidad bastante similares. En el ANEXO VI es posible apreciar que estas entidades son “espol.edu.ec”, “uotavalo.edu.ec”, “utm.edu.ec” y “utmachala.edu.ec”. Todas cuentan con la misma política de privacidad, salvo por las secciones en las cuales se presenta el nombre de la entidad. Si bien las políticas son las mismas (ANEXO VII) el porcentaje de similitud es variante debido al “ruido” presente en cada una de ellas.

En la categoría *government* (*gobierno*), las políticas de privacidad de las entidades tienen una similitud muy alta, salvo el caso de algunas entidades autónomas (gobiernos locales, IESS). Por otra parte, las políticas de privacidad de los ministerios que dependen directamente del Gobierno Central son muy similares, tal como se aprecia en la Figura 3.4.

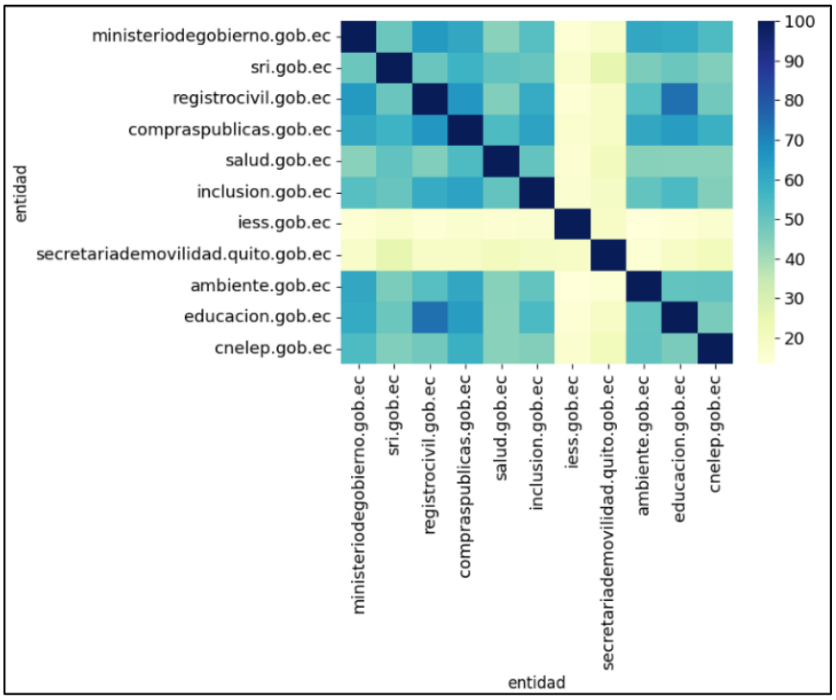


Figura 3.4. Similitud entre políticas de la categoría *Government*

Esta similitud puede deberse a que se han elaborado a partir de una estructura básica recomendada por el Gobierno Central. Es interesante notar que, pese a su similitud, estas políticas cuentan con características únicas de acuerdo con la realidad de cada entidad.

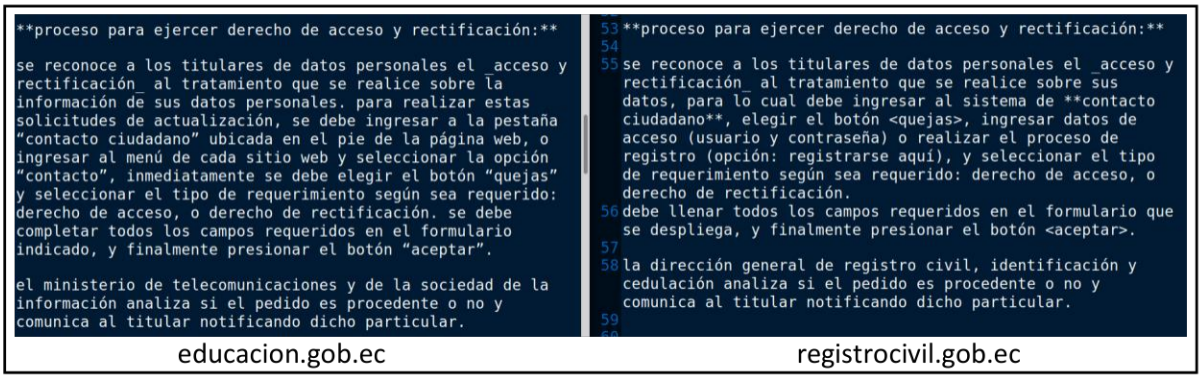


Figura 3.5. Similitud entre "educacion.gob.ec" y "registrocivil.gob.ec" (generalidades)

En la Figura 3.5 se aprecia que el marco general de las políticas de privacidad es el mismo en las entidades gubernamentales. En la figura, esto se evidencia para “educación.gob.ec” y “registrocivil.gob.ec” cuyas políticas están entre las de mayor similitud.

Figura 3.6. Diferencia entre “educacion.gob.ec” y “registrocivil.gob.ec”

En la Figura 3.6 se observa cómo, pese a que ciertas políticas son muy similares, las de entidades gubernamentales sí incluyen detalles propios de cada entidad, por ejemplo, en cuanto a la información recopilada. Esto indicaría que la política sí refleja las particularidades de cada institución.

En la categoría *health* (salud), las políticas de los sitios no tienen un elevado nivel de similitud. Únicamente dos políticas (las de “fmc.com.ec” y “dialcentro.com.ec”) tienen una similitud mayor al 90%, como se muestra en las figuras del ANEXO VIII y del ANEXO IX. De hecho, las dos políticas de privacidad son las mismas, lo que en este caso tiene algo de sentido porque, al investigarlo, los dos sitios pertenecen a la misma empresa.

Aunque, en varios casos, la similitud de las políticas es de esperarse, es raro que sean exactamente las mismas cuando cada entidad podría tener enfoques distintos frente a la recolección de datos.

3.1.6 FRECUENCIA DE PALABRAS DENTRO DE LAS POLÍTICAS DE PRIVACIDAD

A continuación, en la Figura 3.7, se observa la frecuencia de aparición de diferentes palabras en las 67 políticas de privacidad.

Figura 3.7. Nube de palabras en políticas de privacidad ecuatorianas.

Entre las palabras que destacan en la Figura 3.7 están “datos personales” e “información personal”. Esto es una clara referencia al tipo de datos sobre el que tratan las políticas. Sin embargo, del análisis realizado, más allá de estos términos, en general, las políticas no dan más detalle sobre a qué tipos de datos personales se refiere. El único dato personal específico que tiene relevancia en las políticas es “correo electrónico”, como se puede observar en la figura.

Es interesante también la relevancia de “tratamiento datos”. Esto sugeriría el interés de las entidades de informar sobre la existencia de dicho tratamiento de la información recolectada, aunque lamentablemente no se especifiquen detalles de este tratamiento.

Las palabras “términos condiciones” también aparecen con mucha frecuencia, seguramente porque el documento de política de privacidad se suele confundir, o combinar, con el de términos y condiciones, o condiciones de uso de un sistema.

Con menor frecuencia aparecen las palabras “responsable tratamiento”, “seguridad datos” y “uso cookies”, y podrían ser un indicador del intento de que las políticas cumplan con lo dispuesto por la LOPDP. Particularmente, las palabras “responsable tratamiento” son una referencia muy específica a esta ley y a la obligación de informar de un punto de contacto con la entidad para, por ejemplo, hacer un reclamo. Por otro lado, “seguridad datos” refleja la asociación del concepto de seguridad con el de privacidad, en las políticas, aunque son esencialmente distintos. Las palabras “uso cookies” se refieren a uno de los recursos más relevantes para el rastreo de usuarios en línea y, aunque el usuario final promedio no lo conozca muy bien, su mención en la política es necesaria.

Las palabras “derecho acceso” son también una referencia clave a la LOPDP, y particularmente a uno de los derechos pilares relacionados con la privacidad del usuario.

A partir del análisis de las palabras que aparecen en la Figura 3.7, se puede interpretar que las políticas de privacidad y la información que incluyen hacen muy escuetas referencias a la LOPDP, que mencionan ciertos conceptos clave, pero que son insuficientes para que el usuario ejerza su derecho a ser informado sobre lo que se hace con sus datos personales.

3.1.7 ANÁLISIS DE TÓPICOS EN POLÍTICAS DE PRIVACIDAD

Al aplicar modelamiento de tópicos usando LDA, se esperaría identificar grupos de palabras relacionadas con distintos elementos que deberían contener las políticas; por ejemplo, sobre los derechos de los usuarios, los tipos de datos recolectados, o lo que se hace con los datos personales. Sin embargo, en el caso ecuatoriano, el resultado recurrente es la existencia de un solo tópico. Dicho de otro modo, la existencia de un solo tópico es el resultado con más coherencia. Es decir, no existe más de un grupo de palabras que permitan formar categorías separadas. La explicación más lógica para esto sería que las políticas de privacidad ecuatorianas no contienen esos elementos en su estructura y son documentos con un enfoque muy genérico. Esto tiene sentido en este escenario donde la regulación de protección de datos personales ha sido muy recientemente aprobada.

En pocas iteraciones del modelamiento de tópicos, la mayor coherencia se obtenía con 3 tópicos, luego del análisis del gráfico coherencia vs. número de tópicos presentado en la Figura 2.14. En la Figura 3.8 se puede observar los grupos de palabras que el modelo asignó a cada tópico.

Se evidencia que los 3 tópicos tienen básicamente las mismas palabras, bastante genéricas en el contexto de una política de privacidad, lo que explica el fenómeno descrito en el primer párrafo. Aunque los tópicos 1 y 2 incluyen las palabras “correo” y “condiciones”, respectivamente, no parece suficiente para caracterizar una categoría en los documentos. Cabe destacar también que el valor de coherencia obtenido para este modelo es inferior que el del modelo que resulta en un solo tópico.

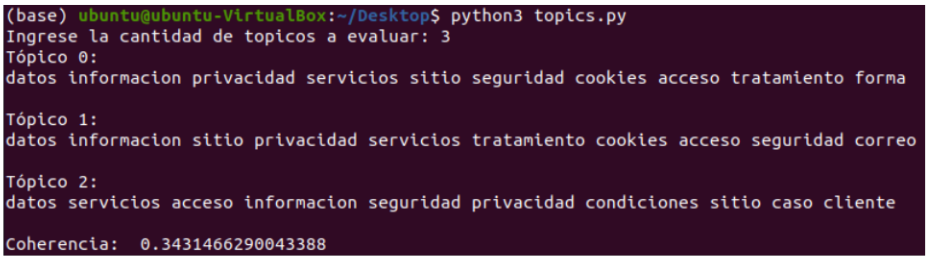


Figura 3.8. Grupos de palabras asociados a cada tópico cuando el modelamiento obtiene una máxima coherencia con 3 tópicos.

3.1.8 CONTRASTE CON MÉXICO

3.1.8.1 Cantidad de palabras y oraciones por política de privacidad

Los sitios mexicanos analizados tienen un promedio de palabras por política de privacidad de aproximadamente 2763 (ANEXO X). A un hispanohablante le tomaría 11 minutos leer cada una de ellas [55]. Tal como con las políticas ecuatorianas (9 minutos), este tiempo parece muy largo, considerando que los usuarios deberían las políticas de todos los sitios que visitan.

En promedio, estas políticas tienen 109 oraciones (ANEXO XI). Esto quiere decir que cada oración tiene, en promedio, 25 palabras. Esta resulta una longitud excesiva [56] y debería acortarse para facilitar la comprensión del documento. Aunque las políticas de privacidad mexicanas son un poco más extensas en promedio que las ecuatorianas, es notorio que las primeras cumplen estrictamente con lo dispuesto en la ley de protección de datos mexicana.

3.1.8.2 Similitud entre políticas de privacidad

Entre las políticas de privacidad mexicanas, solamente 2 tienen una similitud superior al 80% (ANEXO XII), aparentemente debido a que pertenecen a la misma organización (ANEXO XIII). En contraste, 7 políticas ecuatorianas tienen ese nivel de similitud y las entidades involucradas aparentemente no tienen ninguna relación entre sí (sección 3.1.5). Si varias políticas tienen ese nivel de similitud, podría deberse también a que se han reutilizado sin mucho criterio. En todo caso, en México, este fenómeno es prácticamente inexistente.

3.1.8.3 Frecuencia de palabras dentro de las políticas de privacidad

Figura 3.9. Nube de palabras en políticas de privacidad de sitios mexicanos.

En la Figura 3.9 se muestra una nube de palabras generada a partir de las políticas de privacidad mexicanas. Esto, al igual que en las políticas ecuatorianas, es indicador de la principal temática que se aborda en ellas.

Las siguientes palabras más destacadas son “aviso privacidad”, que es el título con el cual se publican las políticas de privacidad en los sitios web. Luego se tiene “tratamiento datos”, que se refiere a la actividad de procesamiento de datos, y que es un término propio de las regulaciones de privacidad.

A diferencia de Ecuador, la descripción de los derechos fundamentales del usuario en torno a la privacidad tiene mayor presencia en las políticas de privacidad mexicanas. Por ejemplo, en estas últimas son relevantes las palabras “acceso”, “cancelación”, “rectificación” y “ejercer derecho”. Adicionalmente, las palabras “derechos arco” que abarcan los derechos de Acceso, Rectificación, Cancelación y Oposición tienen una frecuencia significativa dentro de las políticas de privacidad mexicanas.

La prevalencia de las palabras “ley federal” muestra que en las políticas de privacidad mexicanas se informa de una base legal definida, que es la Ley Federal de Protección de Datos Personales en Posesión de Particulares [9]. En las políticas de privacidad ecuatorianas, en general, no se informa sobre la base legal.

La Figura 3.9 refleja claramente los datos personales que las políticas mexicanas informan que son recolectados. Entre estos datos se encuentran: la “identificación oficial”, “nombre” y “correo electrónico”. En el caso ecuatoriano, solamente aparece el correo electrónico como dato personal recolectado.

México al igual que Ecuador cuenta con un nivel de ruido significativo, que se manifiesta mediante las palabras “sa cv”, “michael page”, “caso”, “grupo financiero” y “clic aquí”. Este ruido podría afectar la comprensión de las políticas de privacidad, ya que son palabras poco relacionadas con este documento, pero que tienen una frecuencia significativa dentro de los documentos analizados.

Con base en las palabras que aparecen en la Figura 3.9, se puede concluir que los documentos analizados se presentan como avisos de privacidad y especifican no solamente los datos personales que son recolectados con el fin de realizar un tratamiento; también informan los derechos de acceso, rectificación y cancelación de los usuarios, y sobre la adecuada aplicación de mecanismos de protección de esos datos personales.

3.1.8.4 Análisis de tópicos en políticas de privacidad

Como resultado del modelamiento de tópicos de las políticas mexicanas, 3 tópicos entregaron una coherencia superior a la obtenida en el caso ecuatoriano ANEXO XIV. Las palabras asociadas a estos tópicos se muestran en la figura del ANEXO XV.

Estas palabras son similares entre los 3 tópicos y están relacionadas con temáticas especializadas en el contexto de una política de privacidad, tal como se observó en la nube de palabras. Entre las palabras que resaltan se encuentran “derechos”, “finalidades”, y “consentimiento” en los tópicos 0 y 1 respectivamente, que representan claramente elementos fundamentales de una política de privacidad. En todo caso, hubiese sido interesante encontrar un tópico relacionado, por ejemplo, con los tipos de datos personales recolectados.

En base a lo expuesto, los tópicos modelados para las políticas de privacidad mexicanas cuentan con términos más especializados en el contexto de una política de privacidad (rectificación, respeto, transparencia, consentimiento, ley), mientras que en Ecuador las políticas de privacidad cuentan con términos más generales.

3.1.9 ANÁLISIS MANUAL

3.1.9.1 Encuesta realizada a usuarios

En el ANEXO XVI es posible apreciar los resultados obtenidos de la encuesta realizada con el fin de contrastar los resultados descritos en las secciones anteriores.

A partir de la encuesta se observa que el tiempo que tardaron los encuestados en leer las políticas de privacidad está entre 5 y 10 minutos. Esto concuerda con los resultados expuestos en la sección 3.1.4.

Adicionalmente, se observa mucha variación en este tiempo dependiendo de la política de privacidad estudiada. En algunos casos la lectura toma menos de 5 minutos y en otros más de 10. Este resultado también se explica por la importante variación en el tamaño de las políticas de privacidad, que se ilustra en el ANEXO II. Es importante recalcar que el porcentaje de políticas de privacidad que han sido leídas en menos de 5 minutos es mayor al porcentaje que tarda más de 10 minutos. Se esperaría que en un futuro las políticas de privacidad requieran menos de 5 minutos para ser leídas y que contengan toda la información planteada en el artículo 12 de la LOPDP.

Un tanto más subjetiva es la pregunta que se refiere a que si el usuario considera que las políticas de privacidad resultan ser largas, en este caso, semejante a la pregunta anterior

más del 21% de políticas de privacidad resultan largas para el usuario lo cual corresponde al 20% de políticas de privacidad que requieren de más de 10 minutos para su lectura.

Es importante considerar que la percepción del usuario es que más del 78% de políticas de privacidad tienen una extensión corta. Aunque esto facilitaría la lectura de estos documentos, podría deberse a que no incluyen toda la información que deberían.

A partir de la encuesta, se advierte que casi la mitad de las políticas de privacidad no incluyen ninguna base legal en la que se fundamenten, aunque antes de la LOPDP, ya existían varios instrumentos legales que garantizaban parcialmente el derecho de los usuarios a la privacidad.

3.1.9.2 Análisis de particularidades

A partir del análisis y procesamiento de las políticas de privacidad ecuatorianas, se destacan a continuación algunos datos interesantes así:

- En la política de privacidad de una de las entidades analizadas (ANEXO XVII) se indica lo siguiente: "[la empresa] se reserva el derecho de vender y distribuir su información personal a otras compañías". Aunque declaraciones en el mismo sentido son comunes en las políticas de privacidad de servicios ofrecidos por grandes empresas, es curioso como en este caso se lo plantea tan directamente.

- Hay un caso particular de política de privacidad de una empresa transnacional que es exactamente la misma que la usada por esta en otro país, en la que no se ha cambiado ni el nombre o dominio de la entidad (ANEXO XVIII). Esto reflejaría el poco interés ciertas entidades por la privacidad de los usuarios ecuatorianos y la necesidad de implementar la obligatoriedad de la LOPDP.

- El ANEXO XIX incluye un par de capturas de políticas de privacidad de entidades como "registrocivil.gob.ec" e "inclusion.gob.ec" que hacen una mención a los datos personales de menores de edad: en el primer caso, para indicar que no se procesan datos de menores, y en el segundo caso para hacer referencia al Código de la Niñez y Adolescencia como base legal para dicho procesamiento.

- Un hecho que es preocupante y se espera que cambie después de la aplicación obligatoria de la LOPDP es que varias entidades tienen una política de uso de cookies en lugar de una política de privacidad (ANEXO XX).

3.1.10 PRÁCTICAS RECOMENDADAS

A continuación, se enumeran algunas prácticas recomendadas para la creación de una política de privacidad, a partir del análisis realizado y los resultados encontrados:

1) Usar como base el artículo 12 de la LOPDP.
2) Permitir el acceso a las políticas de privacidad desde la página principal de la entidad, de otro modo, es más difícil acceder a ella.
3) Publicar la política en el formato más simple y en lenguaje coloquial.
4) Escribir las políticas en oraciones cortas, para así facilitar su comprensión.
5) Limitar la extensión del documento de las políticas de privacidad para que puedan ser leídas en menos de 10 minutos.
6) Excluir términos legales complejos en las políticas de privacidad.
7) Describir la realidad del proceso de extracción y tratamiento de datos realizados por la entidad.
8) Indicar el marco legal en el cual se fundamentan las políticas de privacidad.
9) Especificar claramente cuando se procesará la información de menores.

3.1.11 EVALUACIÓN DE LAS POLÍTICAS DE PRIVACIDAD FRENTE A LAS PRÁCTICAS RECOMENDADAS

En la tabla presente en el ANEXO XXI es posible apreciar la evaluación realizada a las políticas de privacidad, analizadas en función de las prácticas recomendadas. En general, cuatro de cada nueve políticas no cumplen con las prácticas recomendadas. Tres de esas prácticas recomendadas se cumplen parcialmente, y solo dos se cumplen completamente. Cabe anotar que, estas dos últimas prácticas se refieren a la extensión y a la dificultad de lectura de la política de privacidad: ,los usuarios en este caso perciben políticas cortas y no tan complicadas de leer. Sin embargo, como se comentaba previamente, estas características podrían estar relacionadas con documentos incompletos y sin utilidad práctica para el usuario.

A pesar de todo esto, la promulgación de la LOPDP seguramente irá cambiando este escenario y coadyuvando a que la sociedad se comprometa con la privacidad y protección de datos personales.

3.2 CONCLUSIONES

- El nivel de uso de políticas de privacidad en Ecuador es todavía bastante reducido. La aplicación de la LOPDP aún no es obligatoria, pero la política de privacidad es un instrumento básico para el acceso de los ciudadanos a su derecho a la privacidad.

- Las políticas de privacidad no tienen una estructura de redacción estándar, y varía significativamente dependiendo de la entidad. Esto dificulta al usuario la familiarización con estos documentos legales.

- La auditoría automatizada de las políticas de privacidad presenta ciertos obstáculos, particularmente derivados de la falta de estandarización de la página donde se publican. Esto se evidenció en este trabajo, en la dificultad de limpiar el "ruido" al final de la página donde estaban publicadas ciertas políticas.

- La cantidad de oraciones o palabras puede ser un indicador de la facilidad para usar la política por parte de los usuarios, aunque no necesariamente de su calidad en el contexto de la protección de datos.

- El uso de plantillas para generar políticas de privacidad es una opción interesante para ahorrarle trabajo a las entidades, siempre que se adapten para reflejar la realidad de cada organización.

- Las palabras que aparecen con mayor frecuencia en las políticas de privacidad ecuatorianas sugieren que la mayor parte de las políticas de privacidad tienen un enfoque muy general.

- Los tópicos modelados a partir de los documentos analizados permiten inferir que en ellos se incluyen efectivamente conceptos relacionados con privacidad y protección de datos personales.

- Las políticas de privacidad de México debido a la experiencia con la que cuenta el país y a los años que lleva la ley de protección de datos personales vigente, son mucho más informativas y específicas que las políticas de privacidad de Ecuador. Esto garantiza para los usuarios mexicanos un acceso más efectivo a su derecho a la información con mayor facilidad que para los usuarios ecuatorianos.

- Los usuarios perciben que la longitud de las políticas de seguridad en su mayoría es baja, por lo que la apatía que existe en el público general no se debe a su extensión.

- Este trabajo es un insumo fundamental para el estudio posterior de la evolución de las políticas de privacidad en Ecuador a partir de la adopción paulatina de la regulación de protección de datos en el país.

3.3 RECOMENDACIONES

- Los sitios web deberían facilitar el acceso automatizado a sus políticas de privacidad, permitiendo auditorías automáticas de esos documentos a lo largo del tiempo.

- La extensión de las políticas de privacidad no debería exceder las 1500 palabras para asegurar que la mayor parte de usuarios las lean.

- Sería interesante encontrar el balance entre la extensión de la política de privacidad y la información presente en esta política.

- Las políticas de privacidad deberían ser documentos independientes de otros como los relacionados con los términos y condiciones de uso de un servicio. Esto con el fin de facilitar el acceso y la comprensión de la política.

4 REFERENCIAS BIBLIOGRÁFICAS

[1] Naciones Unidas, «La Declaración Universal de Derechos Humanos,» 10 Diciembre 1984.

[2] L. C. Hueso, «Tecnología, libertad y privacidad,» de *Las libertades informativas en el contexto de la web 2.0 y las redes sociales:*, Barcelona, Universidad Autónoma Barcelona, 2018, p. 8.

[3] J. Holvast, «History of privacy,» de *The History of Information Security*, Landsmeer, Privacy Consultants, 2007, p. 30.

[4] Desafíos PWC, «Tecnología por privacidad: ¿qué tan dispuestos están los consumidores a ceder sus datos personales?,» 6 Noviembre 2019. [En línea]. Disponible: https://desafios.pwc.pe/tecnologia-y-privacidad-datos-personales-de-consumidores/. [Último acceso: 16 Diciembre 2021].

[5] G. Paoli, «Gabriela Paoli - Centro de Psicología,» Febrero 2020. [En línea]. Disponible: https://www.gabrielapaoli.com/la-hiperconectividad-influencia-nuestras-vidas/. [Último acceso: 14 Diciembre 2021].

[6] M. Giannechini, P. López, F. Reynoso, B. Rodríguez y E. Velzi, «Ventajas y Desventajas sobre Cloud Computing para las PyMEs en Argentina Rodríguez, Belén- Velzi, Emanuel,» p. 4, 2014.

[7] Asamblea Nacional Constituyente, «Ley Orgánica de Protección de Datos Personales,» *Registro Oficial Suplemento 459,* p. 38, 26 Mayo 2021.

[8] A. Ryan, A. Gunes, L. Elena, K. Mihir, N. Arvind y M. Jonathan, «Privacy Policies over Time: Curation and Analysis of a Million-Document Dataset,» p. 12, 2021.

[9] Cámara de Diputados del Honorable Congreso de la Nación, «Ley Federal de Protección de Datos Personales en Posesión de los Particulares,» p. 18, 5 Julio 2010.

[10] Comisión Europea, «¿Qué son los datos personales?,» [En línea]. Disponible: https://ec.europa.eu/info/law/law-topic/data-protection/reform/what-personal-data_es. [Último acceso: 14 Diciembre 2021].

[11] Real Academia de Lengua Española, «Privacidad,» [En línea]. Disponible: https://dle.rae.es/privacidad. [Último acceso: 14 Diciembre 2021].

[12] Diccionario Panhispánico de español jurídico, «Diccionario Panhispánico de español jurídico,» [En línea]. Disponible: https://dpej.rae.es. [Último acceso: 14 Diciembre 2021].

[13] Jefatura del Estado, «Ley Orgánica de Protección de Datos de Carácter Personal,» 13 Diciembre 1999.

[14] Parlamento Europeo, «Diario Oficial de la Unión Europea,» *Reglamento General de Protección de Datos,* 24 Mayo 2016.

[15] Iclg.com, «USA: Data Protection Laws and Regulations,» 6 Julio 2021. [En línea]. Disponible: https://iclg.com/practice-areas/data-protection-laws-and-regulations/usa. [Último acceso: 16 Diciembre 2021].

[16] Congreso Nacional, «Ley de Comercio Electrónico, Firmas y Mensajes de Datos,» p. 17, 17 Abril 2002.

[17] Asamblea Nacional Constituyente, «Constitución Política de la República de Ecuador de 1998,» 11 Noviembre 1998.

[18] Asamblea Nacional Constituyente, «Ley Orgánica de Telecomunicaciones,» 18 Febrero 2015.

[19] Universidad Tecnológica Latinoamericana, «¿Qué es un aviso de privacidad?,» 15 Junio 2015. [En línea]. Disponible: https://utel.edu.mx/blog/dia-a-dia/que-es-un-aviso-de-privacidad/. [Último acceso: 16 Diciembre 2021].

[20] WebsitePolicies, «What is a Privacy Policy: The Definitive Guide,» [En línea]. Disponible: https://www.websitepolicies.com/blog/what-is-privacy-policy#what-is-a-privacy-policy. [Último acceso: 16 Diciembre 2021].

[21] Cookiebot, «Política de privacidad,» [En línea]. Disponible: https://www.cookiebot.com/es/politica-de-privacidad-para-mi-web/?gclid=CjwKCAiAh_GNBhAHEiwAjOh3ZBXLUxtzzeO2mRKwSsdMyo2z75V0X23fX6o2HO9hGqcGvJo43VEK6xoCWkMQAvD_BwE. [Último acceso: 16 Diciembre 2021].

[22] B. Zhao, «Web Scraping,» *Encyclopedia of Big Data,* p. 3, 2017.

[23] C. Khanna, «Text pre-processing: Stop words removal using different libraries,» Febrero 2010. [En línea]. Disponible: https://towardsdatascience.com/text-pre-processing-stop-words-removal-using-different-libraries-f20bac19929a. [Último acceso: 21 Diciembre 2021].

[24] D. Lee, J. Park, J. Shim y S.-g. Lee, «An Efficient Similarity Join Algorithm with Cosine Similarity Predicate,» *ecture Notes in Computer Science,* vol. 6262, 2010.

[25] S. Wilson, F. Schaub, A. . A. Dara, F. Liu, S. Cherivirala, P. G. Leon, M. S. Andersen, S. Zimmeck, K. M. Sathyendra, N. C. Russell, T. B. Norton, E. Hovy, J. Reidenberg y N. Sadeh, «The Creation and Analysis of a Website Privacy Policy Corpus,» *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics,* p. 11, 2016.

[26] Amazon, «Alexa Topsites,» [En línea]. Disponible: https://www.alexa.com/topsites/countries/EC. [Último acceso: 07 Julio 2021].

[27] Internet Archive, «Wayback Machine,» [En línea]. Disponible: https://archive.org/web/. [Último acceso: 03 Julio 2021].

[28] Archive-it, «About Archive-It,» [En línea]. Disponible: https://archive-it.org/blog/learn-more/. [Último acceso: 13 Enero 2022].

[29] Selenium, «Selenium with Python,» [En línea]. Disponible: https://selenium-python.readthedocs.io. [Último acceso: 11 Septiembre 2021].

[30] Whimboo, «geckodriver,» [En línea]. Disponible: geckodriver. [Último acceso: 11 Septiembre 2021].

[31] Nateprewitt, «Requests: HTTP for Humans,» [En línea]. Disponible: https://docs.python-requests.org/en/latest/. [Último acceso: 11 Septiembre 2021].

[32] Pdfminer.six, «Welcome to pdfminer.six's documentation,» [En línea]. Disponible: https://pdfminersix.readthedocs.io/en/latest/index.html. [Último acceso: 15 Septiembre 2021].

[33] programador clic , «Estudio en profundidad del método de Python para analizar y leer el contenido de archivos PDF,» [En línea]. Disponible: https://programmerclick.com/article/45061352542/. [Último acceso: 31 Septiembre 2021].

[34] Juanpch, «Aviso legal y Política de privacidad,» [En línea]. Disponible: http://www.forosecuador.ec/forum/comunidad/foro-libre/19461-aviso-legal-y-política-de-privacidad. [Último acceso: 15 Enero 2022].

[35] JaimeRamirez-coder, «ANALISIS-AUTOMATIZADO-DE-POLITICAS-DE-PRIVACIDAD-EN-ECUADOR-,» [En línea]. Disponible: https://github.com/JaimeRamirez-coder/ANALISIS-AUTOMATIZADO-DE-POLITICAS-DE-PRIVACIDAD-EN-ECUADOR-.git. [Último acceso: 17 Enero 2022].

[36] Pandas, «pandas documentation,» 12 Diciembre 2021. [En línea]. Disponible: https://pandas.pydata.org/docs/. [Último acceso: 10 Enero 2022].

[37] Pandas, «pandas.DataFrame,» [En línea]. Disponible: https://pandas.pydata.org/docs/reference/api/pandas.DataFrame.html. [Último acceso: 15 Enero 2022].

[38] Pandas, «pandas.DataFrame.to_csv,» [En línea]. Disponible: https://pandas.pydata.org/docs/reference/api/pandas.DataFrame.to_csv.html. [Último acceso: 10 Enero 2022].

[39] Pandas, [En línea]. Disponible: https://pandas.pydata.org/docs/reference/api/pandas.read_csv.html. [Último acceso: 10 Enero 2022].

[40] NLTK Project, «Natural Language Toolkit,» 28 Diciembre 2021. [En línea]. Disponible: https://www.nltk.org. [Último acceso: 17 Enero 2022].

[41] Python Class, «Cosine similarity python sklearn example,» [En línea]. Disponible: https://pythonclass.in/cosine-similarity-python-sklearn-example.php. [Último acceso: 13 Septiembre 2021].

[42] S. Prabhakaran, «Cosine Similarity – Understanding the math and how it works (with python codes),» 22 Octubre 2018. [En línea]. Disponible:

https://www.machinelearningplus.com/nlp/cosine-similarity/. [Último acceso: 30 Septiembre 2021].

[43] Seaborn, «seaborn: statistical data visualization,» [En línea]. Disponible: https://seaborn.pydata.org. [Último acceso: 26 Octubre 2021].

[44] D. Justo, «Las palabras más largas del diccionario,» 17 Mayo 2019. [En línea]. Disponible: https://cadenaser.com/ser/2019/05/17/cultura/1558072790_497530.html. [Último acceso: 20 Enero 2022].

[45] A. Mueller, «WordCloud for Python documentation,» [En línea]. Disponible: https://amueller.github.io/word_cloud/. [Último acceso: 30 Diciembre 2021].

[46] Dataencode, «Topic Modeling in Python : Using Latent Dirichlet Allocation (LDA),» 28 Julio 2020. [En línea]. Disponible: https://dataencode.in/topic-modeling-lda/. [Último acceso: 30 Diciembre 2021].

[47] R. Rehurek, «gensim 4.1.2,» 16 Septiembre 2021. [En línea]. Disponible: https://pypi.org/project/gensim/. [Último acceso: 30 Diciembre 2022].

[48] IBM Watson Explorer, «Lematización,» [En línea]. Disponible: https://www.ibm.com/docs/es/watson-explorer/12.0.x?topic=bases-stemming. [Último acceso: 23 Enero 2022].

[49] Gensim, «Phrase (collocation) detection,» [En línea]. Disponible: https://radimrehurek.com/gensim/models/phrases.html. [Último acceso: 15 Enero 2022].

[50] NLP APIs, «gensim.corpora.Dictionary.compactify,» 2016. [En línea]. Disponible: https://tedboy.github.io/nlps/generated/generated/gensim.corpora.Dictionary.compactify.html. [Último acceso: 23 Enero 2022].

[51] NLP APIs, «gensim.corpora.Dictionary.filter_extremes,» 2016. [En línea]. Disponible: https://tedboy.github.io/nlps/generated/generated/gensim.corpora.Dictionary.filter_extremes.html. [Último acceso: 23 Enero 2022].

[52] El mundo de los datos, «Introducción al topic modeling con Gensim (II): asignación de tópicos,» 31 Marzo 2021. [En línea]. Disponible: https://elmundodelosdatos.com/topic-modeling-gensim-asignacion-topicos/. [Último acceso: 24 Enero 2022].

[53] B. Gastón y L. Juan, «Tutorial #3 Modelado de topicos,» 14 Junio 2021. [En línea]. Disponible: https://bookdown.org/gaston_becerra/curso-intro-r/modelado-de-topicos.html. [Último acceso: 23 Enero 2022].

[54] B. Mabey, «pyLDAvis,» 2015. [En línea]. Disponible: https://pyldavis.readthedocs.io/en/latest/readme.html. [Último acceso: 24 Enero 2022].

[55] Universidad de Extremadura, «Técnicas de estudio: La velocidad lectora,» 7 Septiembre 2021. [En línea]. Disponible: https://biblioguias.unex.es/c.php?g=572102&p=3944889. [Último acceso: 31 Enero 2022].

[56] Ediciones Digitales, «Longitud de oraciones y parrafos,» [En línea]. Disponible: http://edicionesdigitales.info/Manual/Manual/longoracpar.html. [Último acceso: 31 Enero 2022].

5 ANEXOS

ANEXO I. Porcentaje de palabras eliminadas de las políticas de privacidad tras aplicar el filtro automatizado

ANEXO II. Número de palabras por política de privacidad

ANEXO III. Número de oraciones por política de privacidad

ANEXO IV. Similitud de políticas de privacidad de “plusvalía.com” y “multitrabajos.com”

ANEXO V. Similitud de las políticas de privacidad de “puce.edu.ec” y “hospitalvozandes.com”

ANEXO VI. Similitud entre políticas de privacidad de la categoría education (educación)

ANEXO VII. Similitud de las políticas de privacidad de “utmachala.edu.ec”, “utm.edu.ec”, “espol.edu.ec” y “uotavalo.edu.ec”

ANEXO VIII. Similitud entre políticas de la categoría health (salud)

ANEXO IX. Similitud de las políticas de privacidad de “fmc.com.ec” y “dialcentro.com.ec”

ANEXO X. Número de palabras por política de privacidad mexicanas

ANEXO XI. Número de oraciones por política de privacidad mexicanas

ANEXO XII. Similitud mayor al 30% en políticas de privacidad mexicanas

ANEXO XIII. Similitud de las políticas de privacidad de “lasestrellas.tv” y “tudn.mx”

ANEXO XIV. Coherencia vs. Número de tópicos en políticas de privacidad mexicanas

ANEXO XV. Tópicos generados en políticas de privacidad mexicanas

ANEXO XVI. Resultados obtenidos de la encuesta aplicada

ANEXO XVII. Política de privacidad de “expresso.ec”

ANEXO XVIII. Política de privacidad de “claro.com.ec”

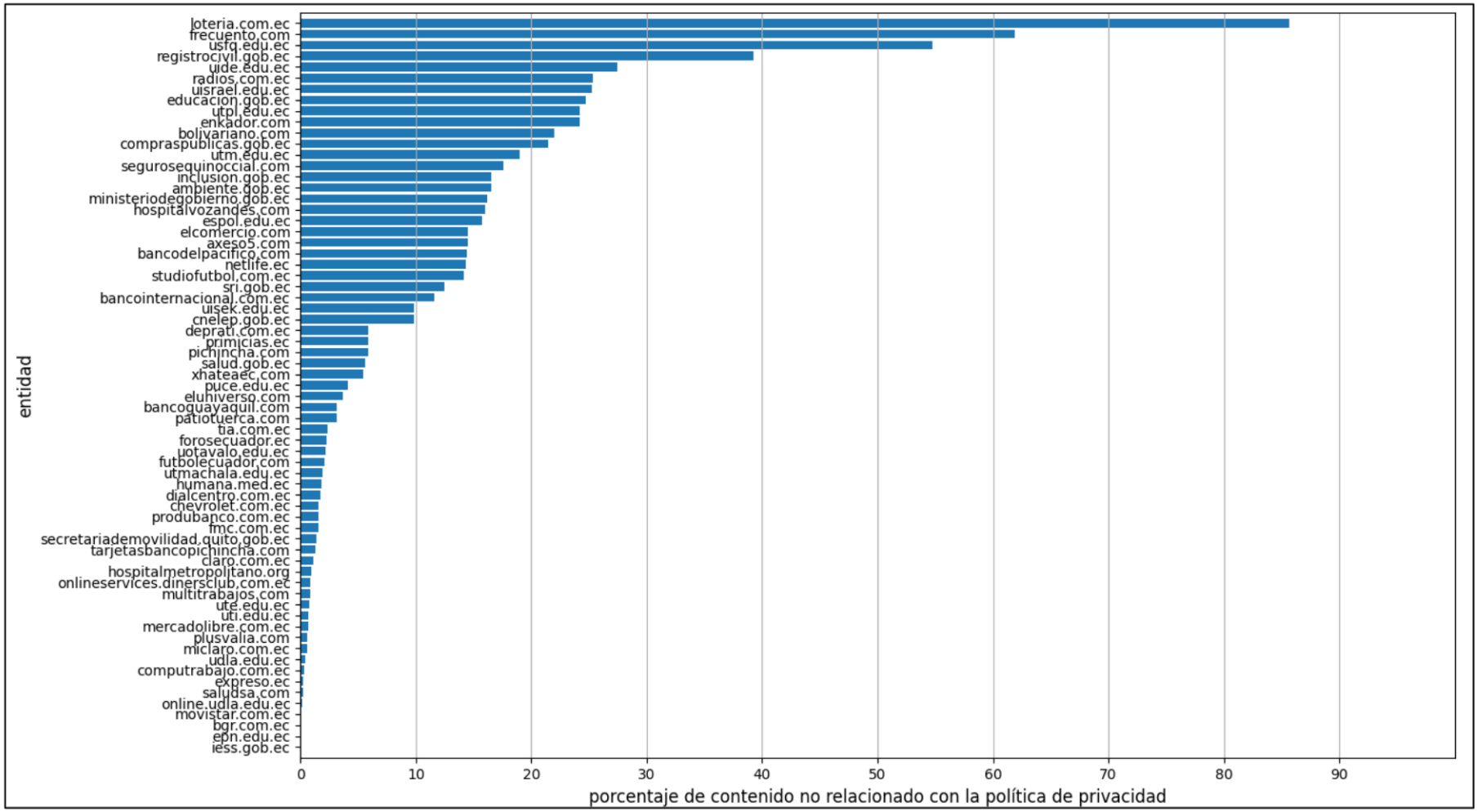
ANEXO XIX. Tratamiento de datos de menores de edad

ANEXO XX. Política de privacidad de “loteria.com.ec”

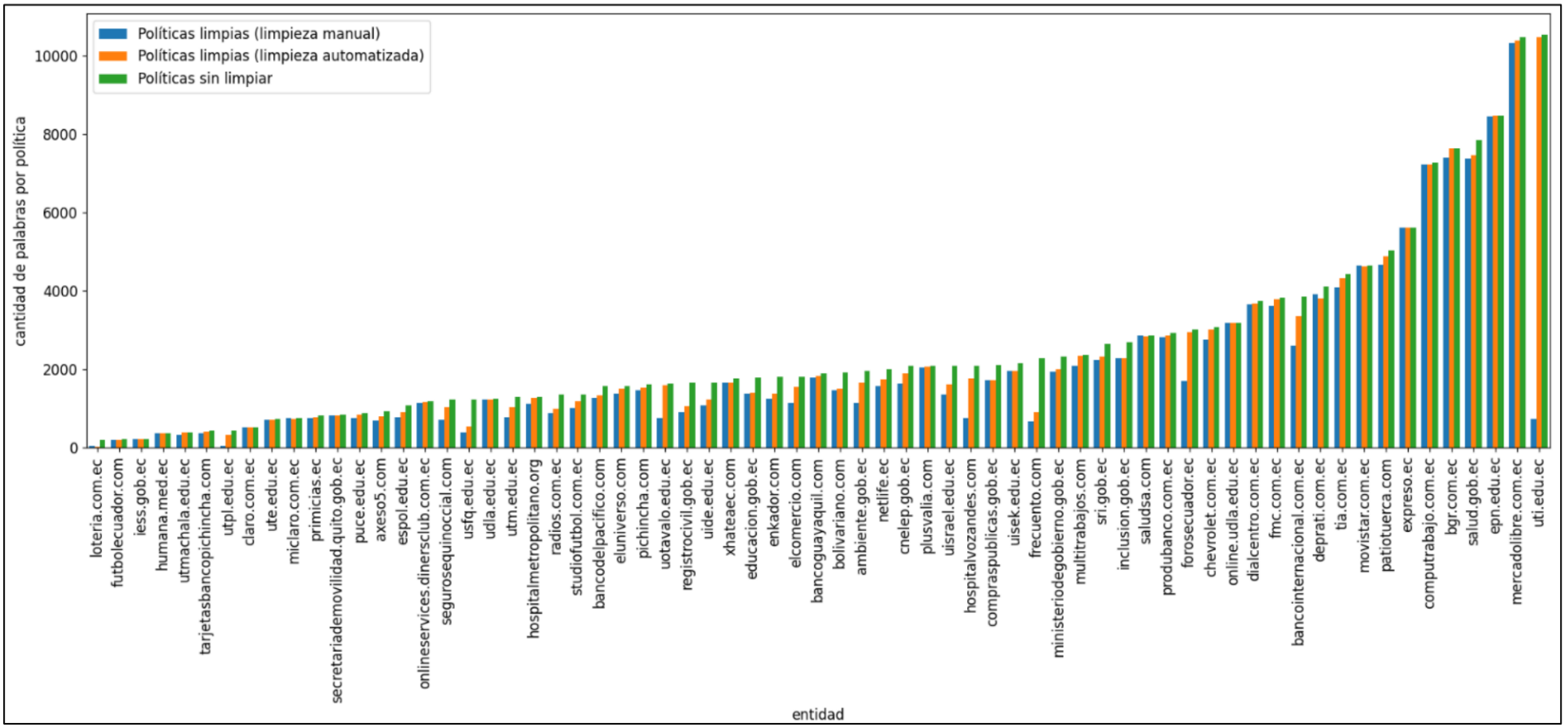
ANEXO XXI. Tabla de evaluación de prácticas recomendadas

ANEXO XXII. Código generado en lenguaje de programación python

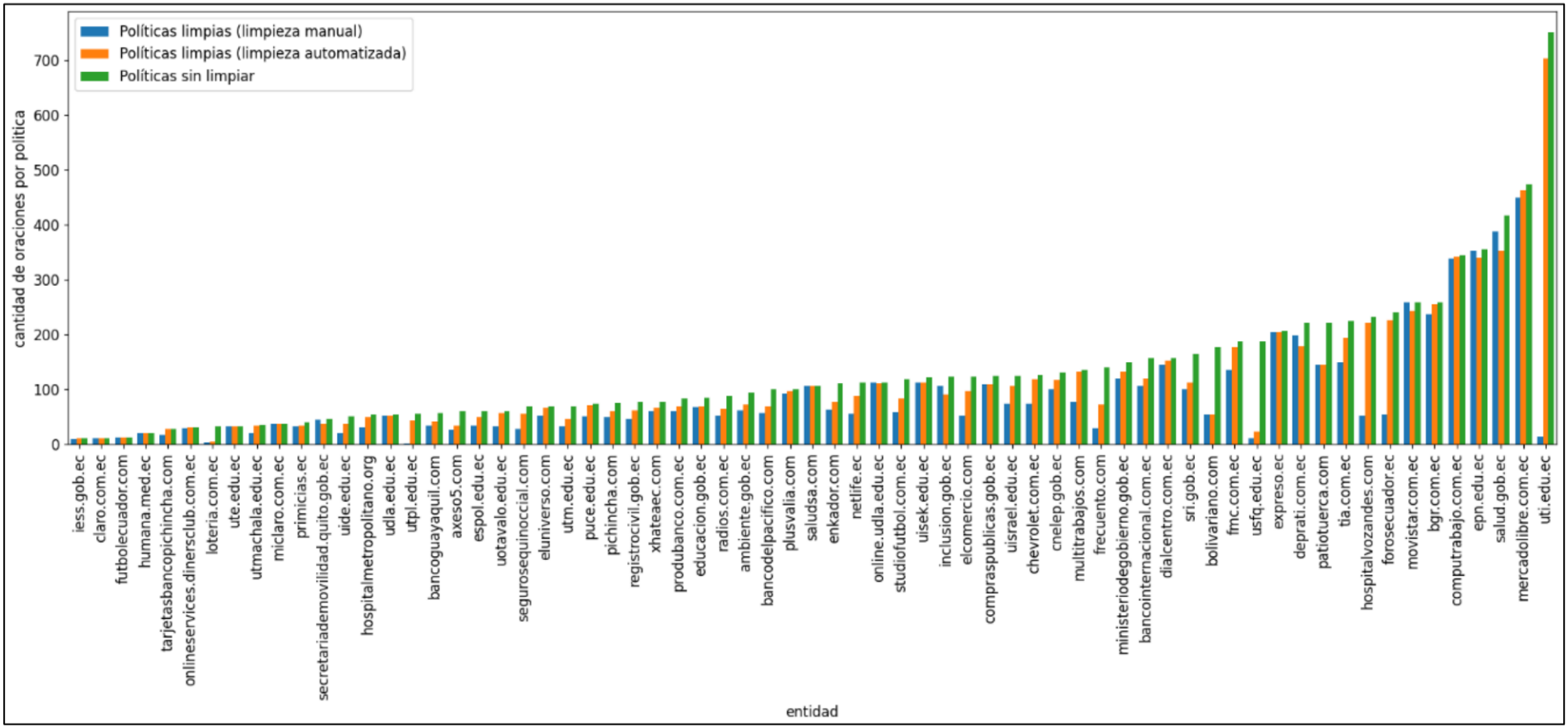
ANEXO I. Porcentaje de palabras eliminadas de las políticas de privacidad tras aplicar el filtro automatizado



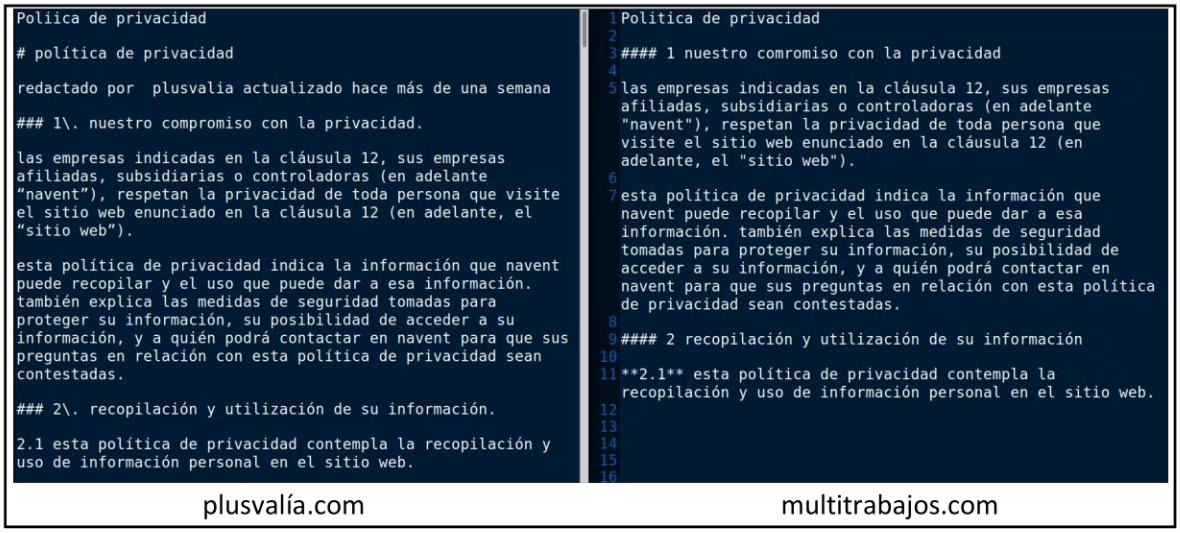
ANEXO II. Número de palabras por política de privacidad



ANEXO III. Número de oraciones por política de privacidad



ANEXO IV. Similitud de políticas de privacidad de “plusvalía.com” y “multitrabajos.com”



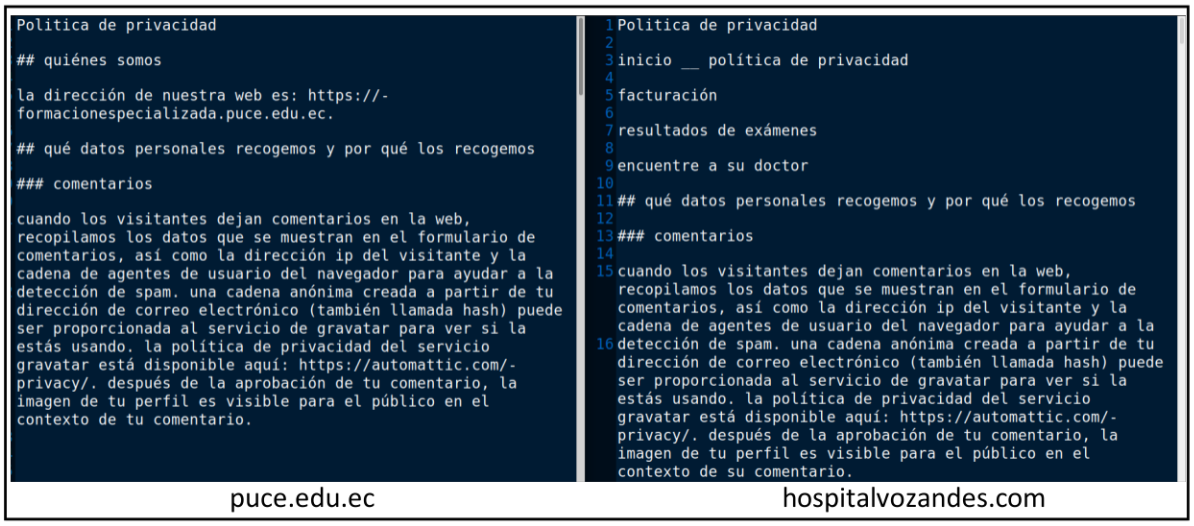
```
Poliica de privacidad

# política de privacidad

redactado por  plusvalia actualizado hace más de una semana

### 1\. nuestro compromiso con la privacidad.

las empresas indicadas en la cláusula 12, sus empresas
afiliadas, subsidiarias o controladoras (en adelante
“navent”), respetan la privacidad de toda persona que visite
el sitio web enunciado en la cláusula 12 (en adelante, el
“sitio web”).

esta política de privacidad indica la información que navent
puede recopilar y el uso que puede dar a esa información.
también explica las medidas de seguridad tomadas para
proteger su información, su posibilidad de acceder a su
información, y a quién podrá contactar en navent para que sus
preguntas en relación con esta política de privacidad sean
contestadas.

### 2\. recopilación y utilización de su información.

2.1 esta política de privacidad contempla la recopilación y
uso de información personal en el sitio web.
```

plusvalía.com

```
Politica de privacidad

#### 1 nuestro comromiso con la privacidad

las empresas indicadas en la cláusula 12, sus empresas
afiliadas, subsidiarias o controladoras (en adelante
"navent"), respetan la privacidad de toda persona que
visite el sitio web enunciado en la cláusula 12 (en
adelante, el "sitio web").

esta política de privacidad indica la información que
navent puede recopilar y el uso que puede dar a esa
información. también explica las medidas de seguridad
tomadas para proteger su información, su posibilidad de
acceder a su información, y a quién podrá contactar en
navent para que sus preguntas en relación con esta política
de privacidad sean contestadas.

#### 2 recopilación y utilización de su información

**2.1** esta política de privacidad contempla la
recopilación y uso de información personal en el sitio web.
```

multitrabajos.com

ANEXO V. Similitud de las políticas de privacidad de “puce.edu.ec” y “hospitalvozandes.com”



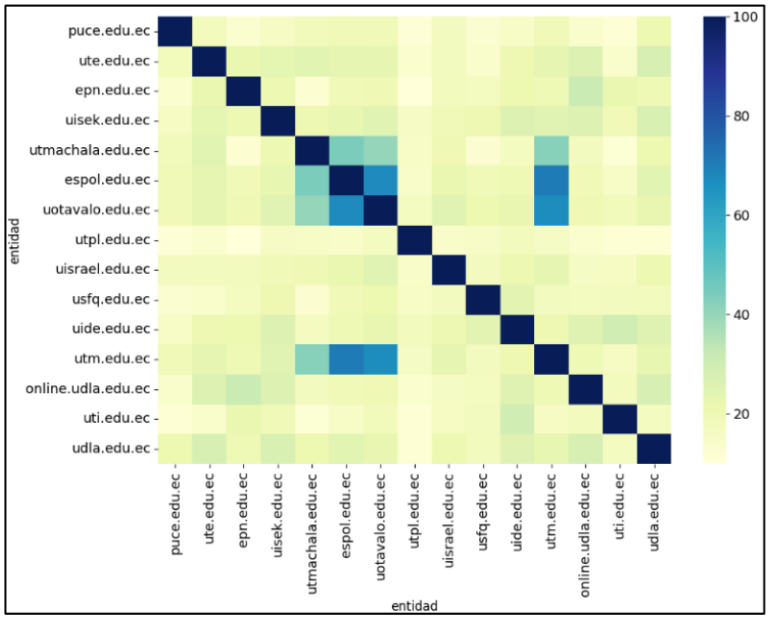
```
Política de privacidad

## quiénes somos

la dirección de nuestra web es: https://-
formacionespecializada.puce.edu.ec.

## qué datos personales recogemos y por qué los recogemos

### comentarios

cuando los visitantes dejan comentarios en la web,
recopilamos los datos que se muestran en el formulario de
comentarios, así como la dirección ip del visitante y la
cadena de agentes de usuario del navegador para ayudar a la
detección de spam. una cadena anónima creada a partir de tu
dirección de correo electrónico (también llamada hash) puede
ser proporcionada al servicio de gravatar para ver si la
estás usando. la política de privacidad del servicio
gravatar está disponible aquí: https://automattic.com/-
privacy/. después de la aprobación de tu comentario, la
imagen de tu perfil es visible para el público en el
contexto de tu comentario.
```

puce.edu.ec

```
Politica de privacidad

inicio __ política de privacidad

facturación

resultados de exámenes

encuentre a su doctor

## qué datos personales recogemos y por qué los recogemos

### comentarios

cuando los visitantes dejan comentarios en la web,
recopilamos los datos que se muestran en el formulario de
comentarios, así como la dirección ip del visitante y la
cadena de agentes de usuario del navegador para ayudar a la
detección de spam. una cadena anónima creada a partir de tu
dirección de correo electrónico (también llamada hash) puede
ser proporcionada al servicio de gravatar para ver si la
estás usando. la política de privacidad del servicio
gravatar está disponible aquí: https://automattic.com/-
privacy/. después de la aprobación de tu comentario, la
imagen de tu perfil es visible para el público en el
contexto de su comentario.
```

hospitalvozandes.com

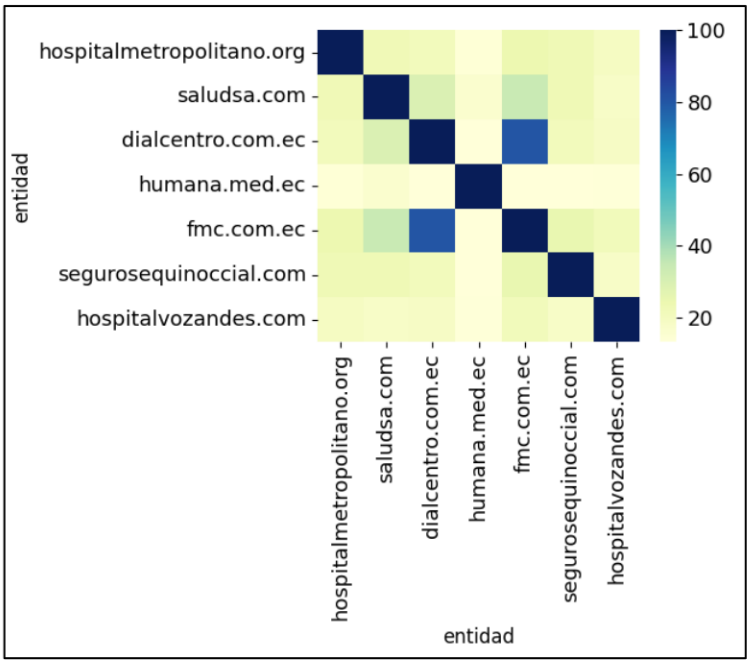
ANEXO VI. Similitud entre políticas de privacidad de la categoría Education (educación)



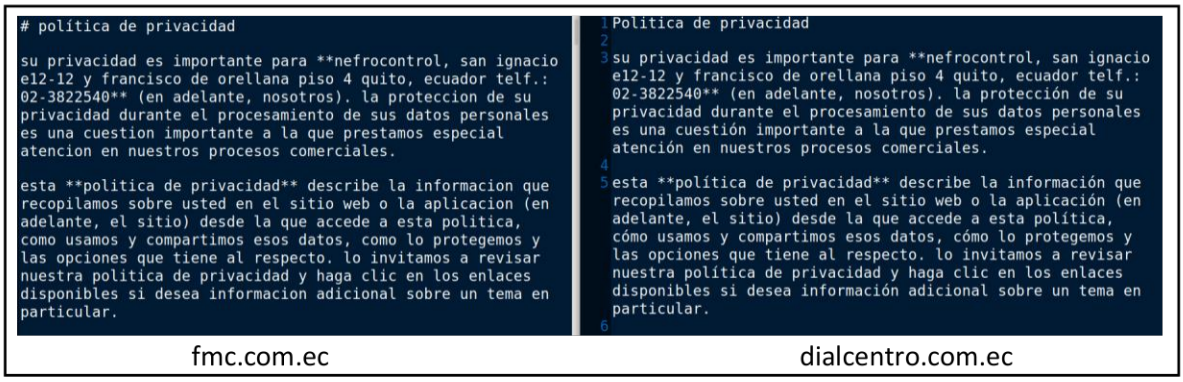
ANEXO VII. Similitud de las políticas de privacidad de “utmachala.edu.ec”, “utm.edu.ec”, “espol.edu.ec” y “uotavalo.edu.ec”

utmachala.edu.ec	utm.edu.ec	espol.edu.ec	uotavalo.edu.ec
#### políticas de privacidad el presente política de privacidad establece los términos en que usa y protege la información que es proporcionada por sus usuarios al momento de utilizar la aplicación. la editorial utmach está comprometida con la seguridad de los datos de sus usuarios. cuando le pedimos llenar los campos de información personal para el registro, con la cual usted pueda ser identificado, lo hacemos asegurando que sólo se empleará de acuerdo con los términos de este documento. sin embargo esta política de privacidad puede cambiar con el tiempo o ser actualizada por lo que le recomendamos y enfatizamos revisar continuamente esta página para asegurarse que está de acuerdo con dichos cambios.	**política de privacidad** el presente política de privacidad establece los términos en que universidad técnica de manabí usa y protege la información que es proporcionada por sus usuarios al momento de utilizar su sitio web. esta compañía está comprometida con la seguridad de los datos de sus usuarios. cuando le pedimos llenar los campos de información personal con la cual usted pueda ser identificado, lo hacemos asegurando que sólo se empleará de acuerdo con los términos de este documento. sin embargo esta política de privacidad puede cambiar con el tiempo o ser actualizada por lo que le recomendamos y enfatizamos revisar continuamente esta página para asegurarse que está de acuerdo con dichos cambios.	# política de privacidad el presente política de privacidad establece los términos en que escuela superior politécnica del litoral usa y protege la información que es proporcionada por sus usuarios al momento de utilizar su sitio web. esta compañía está comprometida con la seguridad de los datos de sus usuarios. cuando le pedimos llenar los campos de información personal con la cual usted pueda ser identificado, lo hacemos asegurando que sólo se empleará de acuerdo con los términos de este documento. sin embargo, esta política de privacidad puede cambiar con el tiempo o ser actualizada por lo que le recomendamos y enfatizamos revisar continuamente esta página para asegurarse que está de acuerdo con dichos cambios.	# política de privacidad el presente política de privacidad establece los términos en que universidad de otavalo usa y protege la información que es proporcionada por sus usuarios al momento de utilizar su sitio web. esta compañía está comprometida con la seguridad de los datos de sus usuarios. cuando le pedimos llenar los campos de información personal con la cual usted pueda ser identificado, lo hacemos asegurando que sólo se empleará de acuerdo con los términos de este documento. sin embargo esta política de privacidad puede cambiar con el tiempo o ser actualizada por lo que le recomendamos y enfatizamos revisar continuamente esta página para asegurarse que está de acuerdo con dichos cambios.

ANEXO VIII. Similitud entre políticas de la categoría Health (salud)



ANEXO IX. Similitud de las políticas de privacidad de “fmc.com.ec” y “Dialcentro.com.ec”



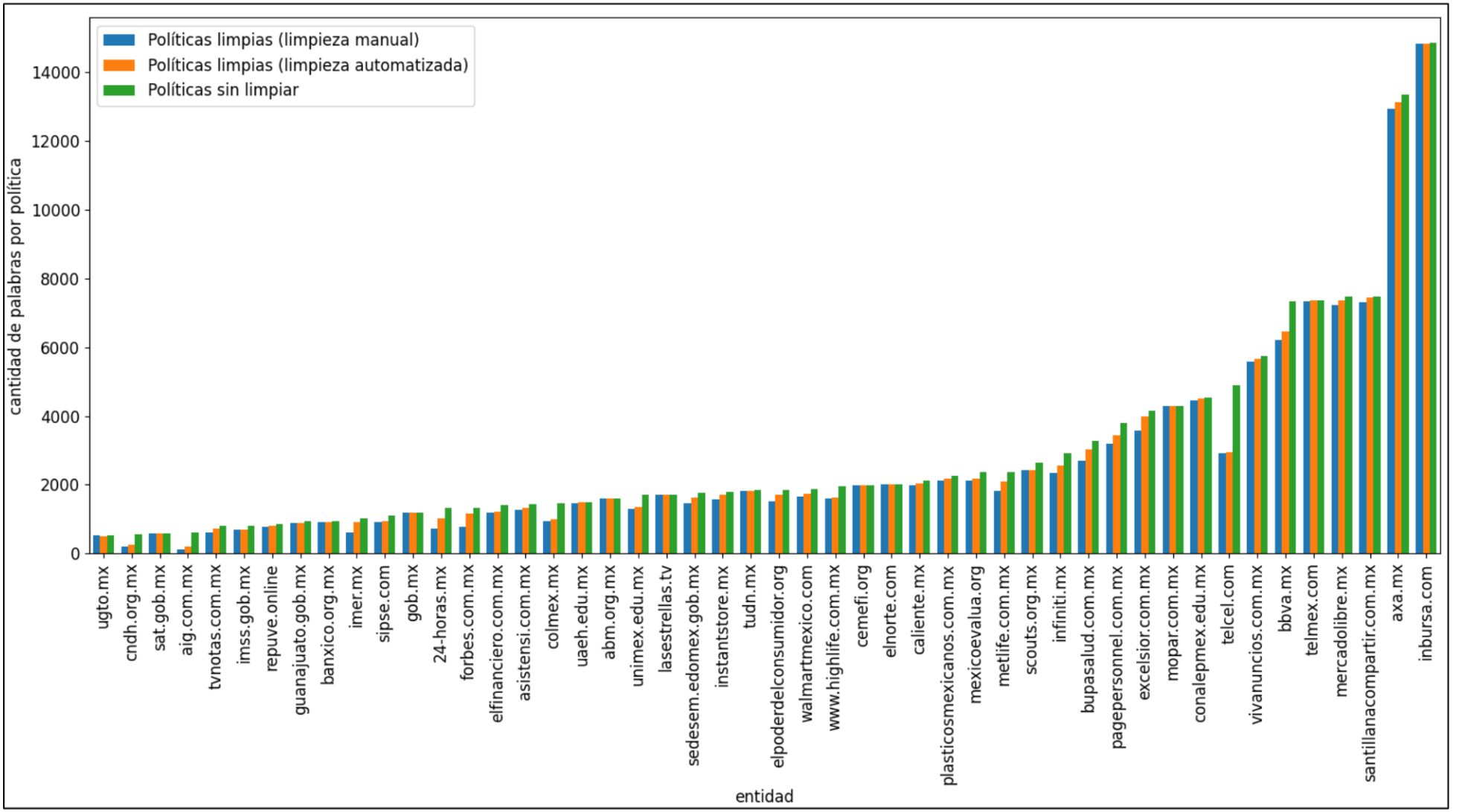
```
# política de privacidad

su privacidad es importante para **nefrocontrol, san ignacio
e12-12 y francisco de orellana piso 4 quito, ecuador telf.:
02-3822540** (en adelante, nosotros). la proteccion de su
privacidad durante el procesamiento de sus datos personales
es una cuestion importante a la que prestamos especial
atencion en nuestros procesos comerciales.

esta **politica de privacidad** describe la informacion que
recopilamos sobre usted en el sitio web o la aplicacion (en
adelante, el sitio) desde la que accede a esta politica,
como usamos y compartimos esos datos, como lo protegemos y
las opciones que tiene al respecto. lo invitamos a revisar
nuestra politica de privacidad y haga clic en los enlaces
disponibles si desea informacion adicional sobre un tema en
particular.
```

fmc.com.ec

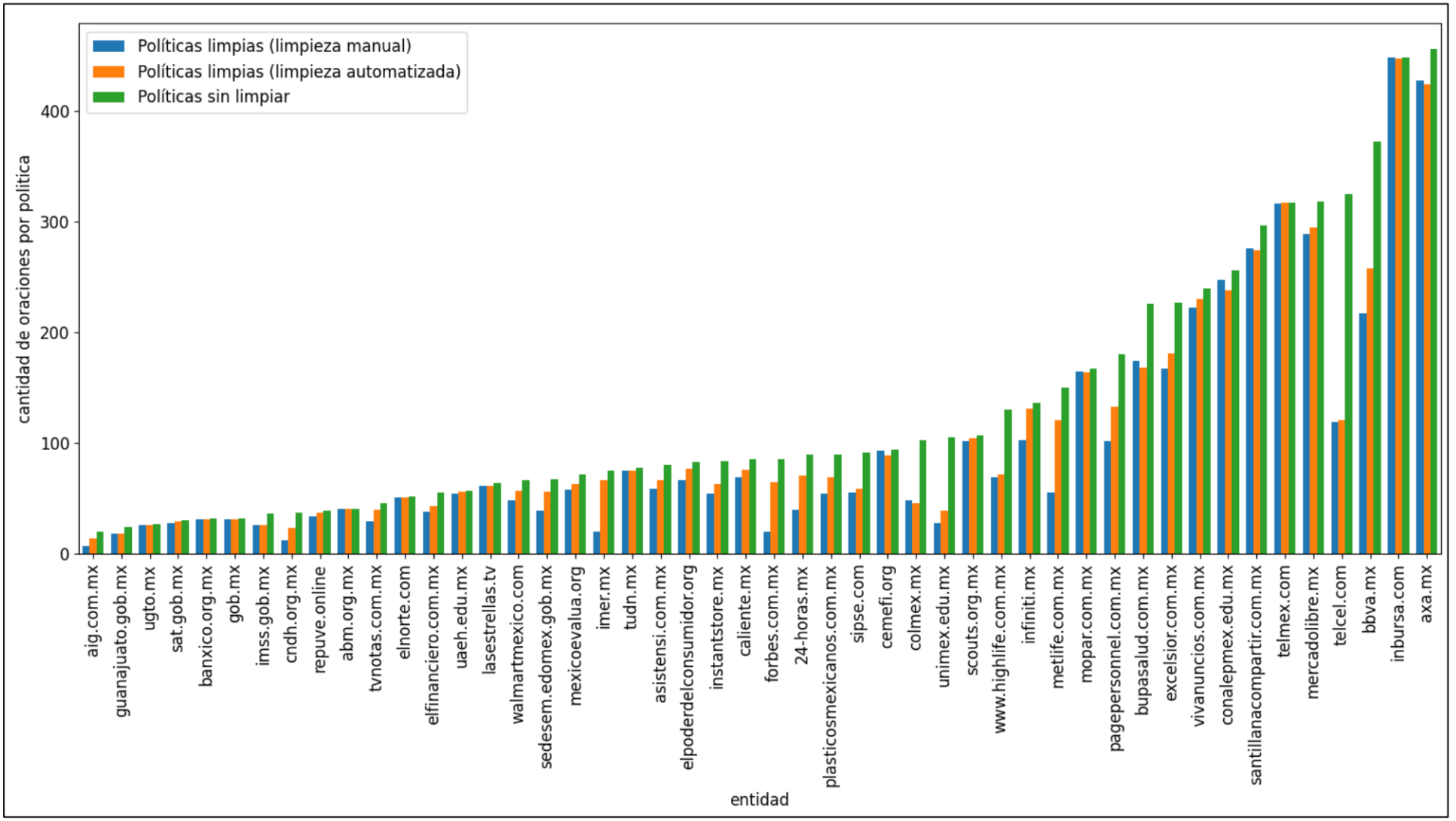
```
Política de privacidad

su privacidad es importante para **nefrocontrol, san ignacio
e12-12 y francisco de orellana piso 4 quito, ecuador telf.:
02-3822540** (en adelante, nosotros). la protección de su
privacidad durante el procesamiento de sus datos personales
es una cuestión importante a la que prestamos especial
atención en nuestros procesos comerciales.

esta **política de privacidad** describe la información que
recopilamos sobre usted en el sitio web o la aplicación (en
adelante, el sitio) desde la que accede a esta política,
cómo usamos y compartimos esos datos, cómo lo protegemos y
las opciones que tiene al respecto. lo invitamos a revisar
nuestra política de privacidad y haga clic en los enlaces
disponibles si desea información adicional sobre un tema en
particular.
```

dialcentro.com.ec

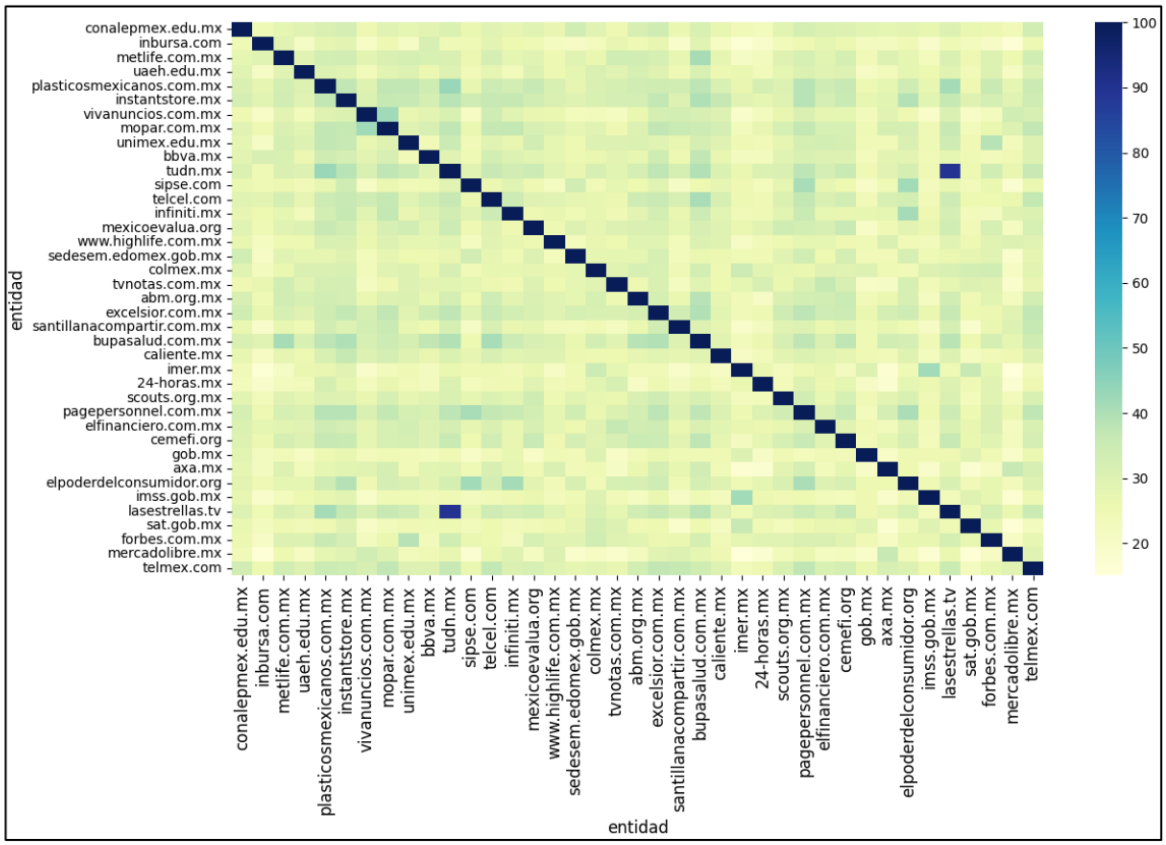
ANEXO X. Número de palabras por política de privacidad mexicanas



ANEXO XI. Número de oraciones por política de privacidad mexicanas



ANEXO XII. Similitud mayor al 30% en políticas de privacidad mexicanas



ANEXO XIII. Similitud de las políticas de privacidad de “lasestrellas.tv” y “tudn.mx”

RESPONSABLE DE LOS DATOS PERSONALES

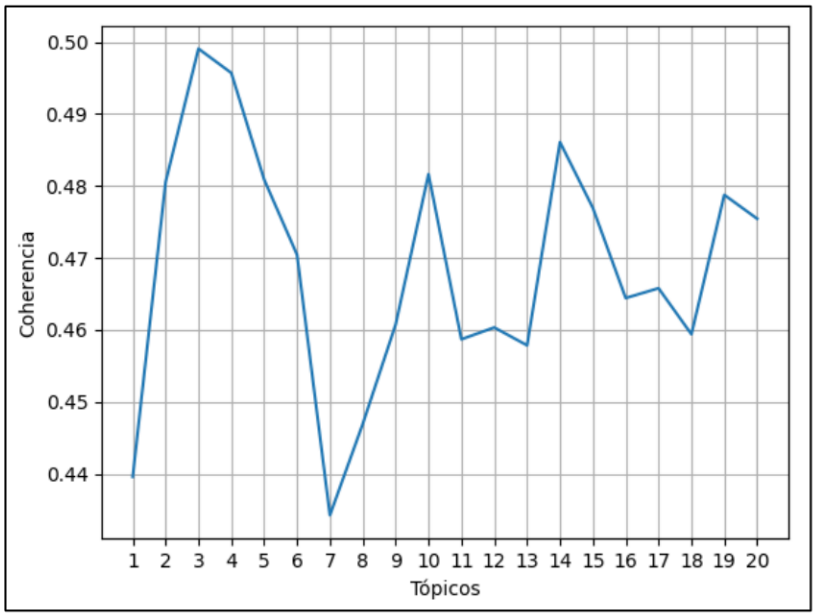
Televisa, S.A. de C.V. (en adelante “Televisa”) con domicilio en Av. Vasco de Quiroga 2000, Colonia Santa Fe, Delegación Álvaro Obregón, C.P. 01210, Ciudad de México, tiene la convicción de proteger la información personal proporcionada por sus usuarios (en adelante “Datos Personales”) y es el responsable de su Tratamiento (término que se define más adelante) cuando sean recabados a través del sitio de Internet https://www.telehit.com/ (en adelante el “Sitio”).

lasestrellas.tv

22 RESPONSABLE DE LOS DATOS PERSONALES
23
24 “Televisa, S.A. de C.V.” (en adelante “Televisa”) con domicilio en Av. Vasco de Quiroga 2000, Colonia Santa Fe, Delegación Álvaro Obregón, C.P. 01210, México, D.F., tiene la convicción de proteger la información personal proporcionada por sus usuarios (en adelante “Datos Personales”) y es el responsable de su Tratamiento (término que se define más adelante) cuando sean recabados a través del sitio de Internet www.televisa.com (en adelante el “Sitio”), medios impresos y/o vía telefónica (en adelante “Otros Medios”).

tudn.mx

ANEXO XIV. Coherencia vs. Número de Tópicos en políticas de privacidad mexicanas

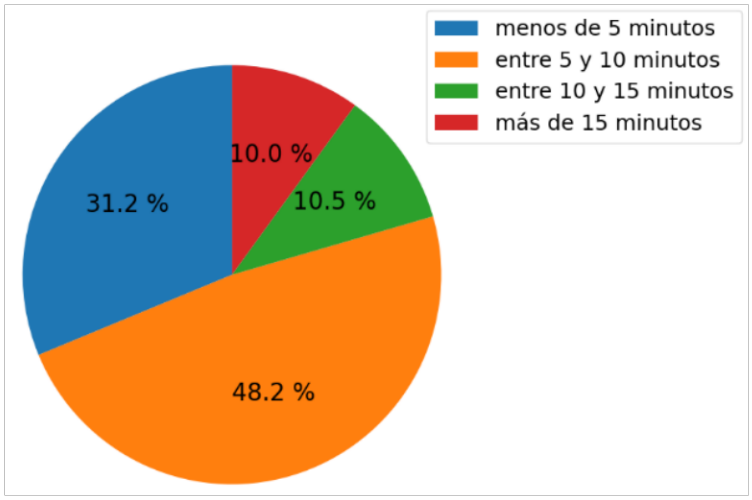


ANEXO XV. Tópicos generados en políticas de privacidad mexicanas

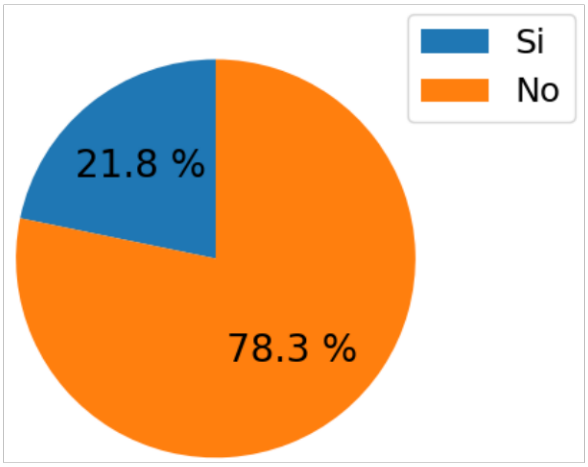
```
(base) ubuntu@ubuntu-VirtualBox:~/Desktop$ python3 topics.py
Ingrese la cantidad de topicos a evaluar: 3
Tópico 0:
aviso tratamiento servicios derechos caso informacion correo finalidades solicitud sitio

Tópico 1:
servicios informacion finalidades caso productos aviso consentimiento finalidad pagina ley

Tópico 2:
aviso servicios caso tratamiento derechos consentimiento informacion ley finalidades solicitud

Coherencia:  0.4990788225091334
```

ANEXO XVI. Resultados obtenidos de la encuesta aplicada

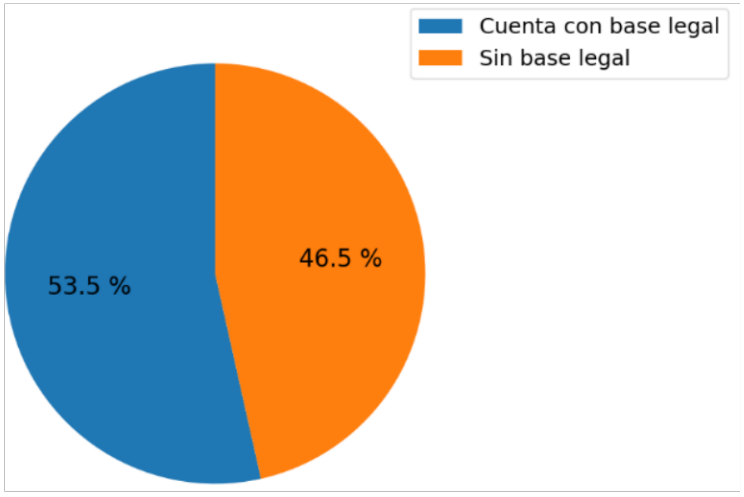
¿Cuánto tiempo le tomó leer la política? (en minutos)



¿La política le pareció muy larga?



¿Cuál es la base legal que se menciona en la política?



ANEXO XVII. Política de privacidad de “expresso.ec”

1\. Información Personal
Existen varias formas dentro de nuestras Páginas en donde usted nos proporcionará datos personales acerca de usted y de sus intereses, tales como nombres, apellidos, género, dirección y dirección de correo electrónico, teléfono convencional y/o celular, profesión u ocupación, datos de facturación, entre otros. Si usted escoge compartir cualquier información con nosotros, ésta será almacenada y usada internamente para su revisión, investigación, análisis, así como otros propósitos de mercadotecnia.
GRANASA se reserva el derecho de vender y distribuir su información personal a otras compañías. Por medio del ingreso de información a este sitio de internet usted acepta que GRANASA pueda vender o distribuir su información personal.
GRANASA divulgará la información personal por usted ingresada en circunstancias especiales, tales como aquellos casos en los cuales, de buena fe considere que la divulgación de la información es razonablemente necesaria para: a) el cumplimiento de procesos legales, b) hacer cumplir los Términos y Condiciones, c) responder a reclamos en los casos que el contenido del servicio viole derechos de terceros, y/o d) proteger los derechos, bienes o seguridad personal de los Usuarios Registrados o del público en general.

ANEXO XVIII. Política de privacidad de “claro.com.ec”

Fecha de recoleccion: 2021-08-26
Politica de privacidad

* CLARO a traves de este acto comunica a todos quienes accedan a su sitio web www.clarochile.cl, la siguiente Politica de Privacidad.
* CLARO solo efectuara tratamiento de datos personales respecto de aquellos que han sido entregados voluntariamente por los clientes y/o usuarios a traves del sitio web www.clarochile.cl
* CLARO establece las debidas politicas de seguridad y controles destinados a velar por la confidencialidad de los datos personales de todos los clientes y/o usuarios que se registren como tales en el sitio web www.clarochile.cl, mediante las secciones y canales establecidos y habilitados para dichos efectos.
* Cabe hacer presente que, en virtud del mandato legal respectivo y en ocasion de un proceso judicial en curso, las Autoridades y/o Instituciones Gubernamentales a cargo de esta gestion, podran solicitar a CLARO, informacion de caracter personal de sus Clientes y/o Usuarios sin consentimiento o autorizacion de estos ultimos.
* Sin perjuicio de lo anterior, Claro Chile, podra discutir, cuestionar y pedir aclaracion del alcance del requerimiento a la autoridad solicitante, con el objeto de resguardar y proteger la privacidad de los datos personales de sus Clientes y/o Usuarios.
* El usuario y/o cliente autoriza a CLARO y sus empresas relacionadas, filiales o matrices de conformidad a la

ANEXO XIX. Tratamiento de datos de menores de edad

```
**Finalidad:** Utilizamos su información para mejorar el contenido, la
usabilidad y experiencia de los usuarios del portal web
www.registrocivil.gob.ec de la DIGERCIC.
_Datos personales de niños:_ Al ser un canal informativo NO recopilamos datos
de o sobre niños.
```

Política de privacidad de “registrocivil.gob.ec”

```
EL MIES no podrá recopilar datos sobre niñas, niños y adolescentes sin previo
consentimiento y autorización de sus representantes legales por considerarse
datos sensibles en estricto cumplimiento a lo establecido en el artículo 44 de
la Constitución de la República del Ecuador y 11 del Código de la Niñez y
Adolescencia; esta Cartera de Estado cuenta con una variedad de sistemas
tecnológicos como Antivirus, Sistemas de Seguridad Perimetral, Monitoreo,
etc., para detectar y abordar actividades anómalas y evitar el uso indebido de
la información.
```

Política de privacidad de “inclusion.gob.ec”

ANEXO XX. Política de privacidad de “loteria.com.ec”

```
Este sitio utiliza cookies, incluidas las cookies de creación de perfiles de
terceros. Si quieres saber más o negar el consentimiento a todas o algunas
cookies, lee la información. Si continúa navegando, acepta el uso de todas las
cookies.
```

ANEXO XXI. Tabla de evaluación de prácticas recomendadas

Prácticas	Evaluación	Comentario	Indicador
1	No cumplen	La mayor parte de políticas no incluye toda la información requerida por el artículo 12 de la LOPDP	
2	Cumplen parcialmente	Un porcentaje considerable de entidades no permite el acceso a la política de privacidad desde su página principal.	
3	No cumplen	Existe un nivel de ruido considerable en la mayor parte de sitios web en los cuales se encuentran las políticas de privacidad.	
4	No cumplen	El promedio de palabras por oración es de 21 palabras, lo cual es excesivo y genera dificultad al leer.	
5	Cumplen	La mayor parte de las políticas tienen una extensión corta en base a la encuesta realizada a los usuarios.	
6	Cumplen	En general las políticas de privacidad son claras y sin términos legales que dificulten su lectura.	
7	Cumplen parcialmente	Existen un porcentaje elevado de políticas de privacidad que son copiadas de otras entidades o que han sido generadas con plantillas por lo cual no representan la realidad de cada entidad	
8	Cumplen parcialmente	Aproximadamente la mitad de las políticas de privacidad no indican la base legal en la cual se fundamentan.	
9	No cumplen	En general, las políticas de privacidad no indican si procesan o no información de menores.	

ANEXO XXII. Código generado en lenguaje de programación Python

En el link adjunto se encuentra el código generado para realizar el análisis automatizado.

https://github.com/JaimeRamirez-coder/ANALISIS-AUTOMATIZADO-DE-POLITICAS-DE-PRIVACIDAD-EN-ECUADOR-/tree/main/Scripts%20Desarrollados