

Memoria de Pasantía de Investigación

entre GobLab UAI y LabHacker



Editan:



P o l i L a b



 **Lab Hacker**



Memoria de Pasantía de Investigación entre GobLab UAI y LabHacker

Andrés Letelier

Director de Investigación, GobLab UAI

andres.letelier@edu.uai.cl

Junio de 2018

Resumen

El presente documento detalla las actividades realizadas durante la visita de Andrés Letelier, Director de Investigación del GobLab UAI, al Lab Hacker de la *Câmara dos Deputados* de Brasil, en el marco de una pasantía de colaboración e intercambio de ideas.

Durante la pasantía se presentaron los proyectos y metodologías de ambos laboratorios. Se incorporaron datos del Senado chileno a *Babel*, herramienta diseñada para la recopilación y disponibilización estructurada de manifestaciones de figuras públicas. Finalmente, se estudió la aplicabilidad de algunas técnicas de procesamiento de lenguaje natural para analizar estos textos.

La pasantía fue financiada por InnoLabs, red de laboratorios de gobierno para la innovación pública en Iberoamérica.

Índice general

1. Introducción	2
2. Organizaciones involucradas	3
2.1 InnoLabs	3
2.2 GobLab UAI	3
2.3 LabHacker	4
3. Proyectos de LabHacker	5
3.1 Retórica Parlamentar	5
3.2 Tenho Dito	5
3.3 Babel	7
3.3.1 Base de datos	7
3.3.2 API	8
3.3.3 Carga de datos y análisis	8
3.4 E-Democracia	8
3.4.1 Audiências Interativas	8
3.4.2 Wikilegis	8
3.4.3 Expressão	10
4. Actividades realizadas	11
4.1 Lunes	11
4.2 Martes	11
4.3 Miércoles	14
4.4 Jueves	14
4.5 Viernes	16
5. Conclusiones	17

1. Introducción

El aumento de las demandas ciudadanas por mayor transparencia y participación en las decisiones públicas, junto a los bajos niveles de confianza en sus representantes, han fomentado el interés y la necesidad de innovación pública.

En respuesta, han surgido nuevas organizaciones “híbridas” que combinan la investigación en políticas públicas con la tecnología y la ciencia de los datos: los laboratorios de innovación pública. Estos laboratorios están dando una inyección de creatividad y experimentación al diseño de políticas y programas públicos, reuniendo diferentes actores del ecosistema público para explorar nuevos enfoques y soluciones a problemas complejos de manera cooperativa. En un documento previo de la Red Innolabs, elaborado por Exequiel Rodríguez, Directora: Rita Grandinetti, PoliLab UNR, se profundiza sobre estos espacios para la innovación pública a partir de la identificación, caracterización y mapeo comparado de las experiencias internacionales. El documento se puede descargar [aquí](#)

Los laboratorios de innovación pública se caracterizan por la búsqueda de soluciones a problemas públicos, a través de diversas tecnologías basadas en la colaboración y participación. Entre ellas el uso intensivo de datos abiertos y la aplicación de tecnologías experimentales. Sin embargo, estas organizaciones se enfrentan a una serie de retos para estimular la innovación en el sector público. Entre éstos, algunos de los principales son: el ser tenidos en cuenta como actores clave por los organismos públicos en temas y procesos centrales de sus agendas; lograr la difusión, adopción efectiva y transferencia de conocimiento de las innovaciones que desarrollan, e impulsar la sostenibilidad de los ecosistemas de innovación pública.

El proyecto de la Red InnoLabs, pretende responder a estos desafíos mediante el intercambio de conocimiento y experiencia entre laboratorios de innovación pública en Iberoamérica, fomentando la creación conjunta de metodologías y procesos de innovación pública para responder a los problemas de la sociedad actual, y creando conciencia en la ciudadanía y los actores del ecosistema público sobre el papel de estos nuevos actores en la creación de políticas públicas.

Es en este contexto que se crea la presente pasantía de investigación, con el fin de promover el intercambio de conocimientos y la colaboración entre laboratorios de investigación de la red. Para esto, se organizó una visita de una semana de Andrés Letelier, director de investigación del GobLab UAI, a LabHacker, de la Cámara dos Diputados de Brasil, con los siguientes objetivos específicos:

- Conocer los proyectos de LabHacker.
- Estudiar el diseño e implementación de Babel (descrito en la sección 3.3), evaluando la posibilidad de incorporar datos chilenos a la base de datos y entendiendo las posibles dificultades asociadas al levantamiento de datos.
- Presentar algunos de los proyectos de investigación del GobLab y el modelo de formación de funcionarios públicos.

Este documento detalla las actividades realizadas entre el 11 y el 15 de junio durante la pasantía y las principales conclusiones obtenidas.

2. Organizaciones involucradas

2.1 InnoLabs

InnoLabs es una red de Laboratorios de Innovación Pública integrada por NovaGov.Lab, PoliLab (Universidad Nacional de Rosario, Argentina), U-GOB Tecnología en Gobierno (México), Laboratorio Hacker de la Cámara de Diputados de Brasil, GobLab de la Universidad Adolfo Ibañez (Chile), y LABCapital (Veeduría Distrital de Bogotá, Colombia), una alianza para impulsar la difusión, adopción efectiva y transferencia de conocimiento de las innovaciones, y promover la sostenibilidad de los ecosistemas de innovación pública.

La iniciativa tiene por objetivos:

1. Ser un canal para compartir visiones, conocimiento y experiencias sobre las prioridades de las agendas institucionales y de actores clave;
2. Convertirse en espacio de debate sobre los laboratorios, para plantear ideas y líneas de trabajo conjuntas de mejora de métodos de trabajo y de enfrentar problemas comunes y los importantes retos de la sociedad;
3. Sensibilizar a los actores del ecosistema público sobre la importancia de participar en los espacios de innovación pública.
4. Concienciar a las administraciones públicas sobre la importancia de los laboratorios en la creación de políticas públicas, así como
5. contribuir a la provisión de asistencia.

2.2 GobLab UAI

El GobLab UAI¹ es el laboratorio de innovación pública de la Escuela de Gobierno de la Universidad Adolfo Ibañez. Fue creado a principios de 2017 y su misión es promover el uso de la ciencia de datos en el sector público, para mejorar la gestión pública y generar políticas públicas basadas en la evidencia.

El GobLab busca aportar a la comunidad a través de tres ramas principales: investigación, docencia y proyectos aplicados de ciencia de datos, con el fin de transformar el sector público y mejorar la vida de las personas.

En el área de docencia, el GobLab ha desarrollado un programa de formación intensiva de analistas de datos para el sector público, que está implementado en un programa de diplomado de dos semestres. Este programa abarca desde la elaboración hasta la implementación de proyectos de datos. Está orientado a profesionales del sector público en roles de análisis, estudios y fiscalización, y a profesionales de centros de estudios, gremios, fundaciones y medios de comunicación dedicados al análisis de datos.

En el área de investigación y desarrollo de proyectos, el GobLab ha diseñado e implementado prototipos de modelos predictivos orientados principalmente a la fiscalización, usando técnicas de *machine learning* para focalizar esfuerzos de prevención.

2.3 LabHacker

LabHacker² es el laboratorio de investigación de la *Câmara dos Deputados* de Brasil. LabHacker es un espacio para promover el desarrollo colaborativo de proyectos innovadores de la ciudadanía relacionados al Poder Legislativo, y se especializa en temas de *crowdlaw* y datos abiertos.

Dentro de la Cámara de Diputados, el laboratorio tiene el objetivo de articular una red entre parlamentarios, hackers y sociedad civil que contribuya a la cultura de la transparencia y de la participación social a través de la gestión de datos públicos.

Además de estimular el desarrollo de aplicaciones digitales que permitan la visualización más intuitiva y simplificada de las informaciones legislativas, LabHacker promueve actividades como maratones hackers: *hackatones* sobre temas innovadores que incluyen talleres de creación y discusiones con especialistas y parlamentarios. Están orientados a estudiantes universitarios y de escuelas públicas, y niños interesados en tecnología.

El laboratorio también es responsable de la gobernanza del portal e-Democracia, una plataforma de participación que amplía la transparencia de la actividad legislativa y la interacción entre los parlamentarios y la sociedad.

Cualquier persona interesada en proyectos de ciudadanía, especialmente programadores y desarrolladores de software, puede trabajar libremente en el ambiente de LabHacker.

¹ <https://gobierno.uai.cl/centro-investigacion/goblab-uai/>

² <http://labhacker.cd.leg.br/>

3. Proyectos de LabHacker

Este capítulo describe algunos de los proyectos de LabHacker estudiados durante la pasantía. Todos estos proyectos son de código abierto y están disponibles en el GitHub de LabHacker³.

3.1 Retórica Parlamentar

Retórica Parlamentar⁴ es una aplicación web que permite descubrir y visualizar los temas más importantes discutidos en el parlamento año a año.

La generación de temáticas está hecha en base a *clustering* con *k-means*⁵ a partir del texto de los discursos parlamentarios. Las temáticas se muestran en una nube de palabras, donde el usuario puede ver la importancia relativa y los parlamentarios que más hablan de cada una de ellas. La figura 3.1 muestra una captura de pantalla de la aplicación con datos del año 2015.

La aplicación fue desarrollada durante la primera *Hackathon da Câmara dos Deputados* por Davi Moreira, Manoel Galdino y Luis Carli. Posteriormente fue incubada por LabHacker. Fue actualizada por última vez en el año 2016, siendo reemplazada por *Tenho Dito*.

3.2 Tenho Dito

Tenho Dito⁶ es la continuación de Retórica Parlamentar. Permite visualizar las temáticas principales que se están discutiendo actualmente por cada estado brasileño. Fue creada por Matheus Fernandes para su trabajo de graduación, con el título *Tenho Dito: uma aplicação para análise de discursos parlamentares utilizando técnicas de processamento de linguagem natural*⁷

La agrupación temática se hace usando *naïve bayes*⁸ clasificando cada discurso en una de 22 temáticas. La figura 3.2 muestra una captura de pantalla de la aplicación para el estado de São Paulo.

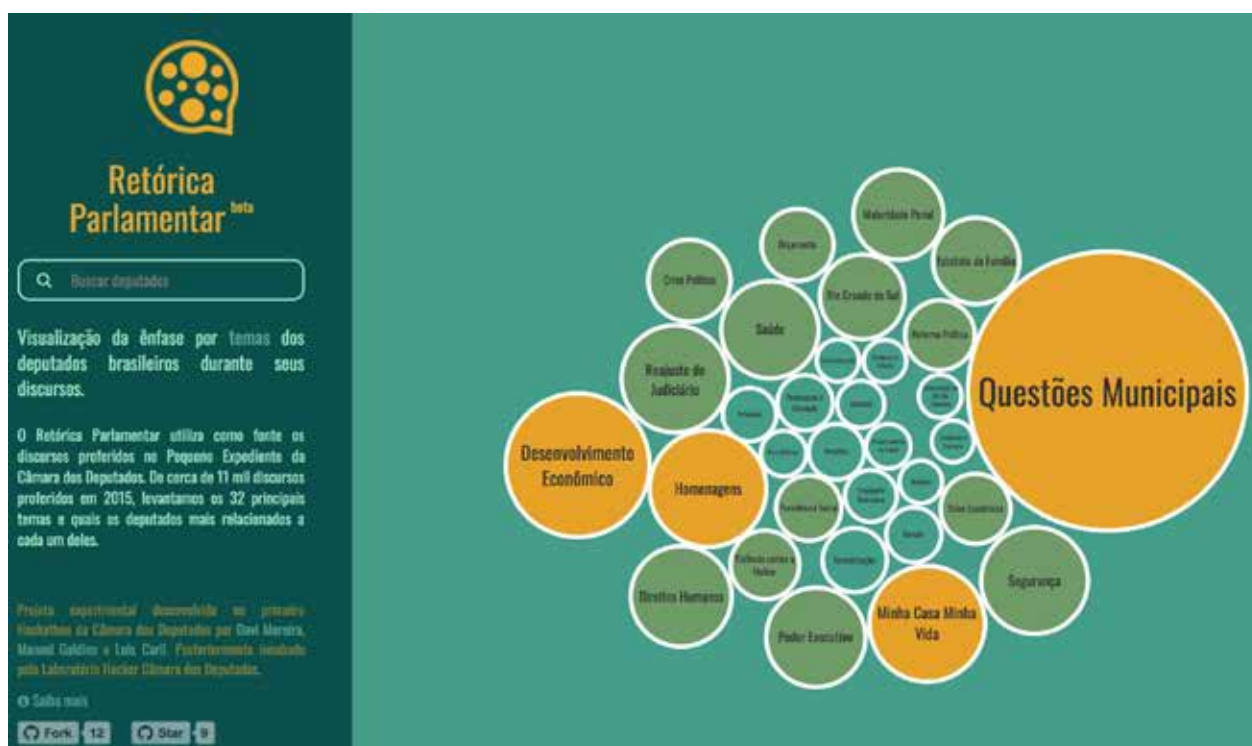


Figura 3.1: Captura de pantalla de Retórica

³ <https://github.com/labhackercd>

⁴ <http://retorica.labhackercd.leg.br>

⁵ <https://es.wikipedia.org/wiki/K-means>

⁶ <http://tenhodito.labhackercd.leg.br>

⁷ <http://bdm.unb.br/handle/10483/19846>

⁸ https://es.wikipedia.org/wiki/Clasificador_bayesiano_ingenuo



Figura 3.2: Captura de pantalla de Tenho Dito

3.3 Babel

Babel⁹ es la continuación de Tenho Dito. A diferencia de las iteraciones anteriores, Babel pretende agregar todo tipo de *manifestación* de figuras públicas: tweets, publicaciones de Facebook, discursos en cámara, proyectos de ley, entrevistas y otros.

A diferencia de los proyectos anteriores, Babel quiere ser un repositorio que permita centralizar datos levantados de manera distribuida, además de una fuente de datos estructurados para facilitar el análisis de investigadores. Con esto se busca facilitar la colaboración de analistas, permitiendo compartir los datos que se levanten como parte de un proyecto de análisis de texto de figuras públicas para que otros investigadores puedan realizar sus propios estudios.

Babel está compuesto por tres subproyectos: una base de datos, una API y una aplicación de análisis. Estas tres partes están construidas como una aplicación web en django,⁸ un *framework* de python para desarrollo web.

3.3.1 Base de datos

Babel tiene un modelo de datos que considera que cada figura pública (por ahora solamente parlamentarios) emiten manifestaciones por distintos canales de comunicación: Twitter, Facebook, la prensa, los discursos, proyectos de ley, etcétera. El considerar esto permite agrupar fácilmente los distintos discursos públicos por su autor sin importar el canal por el que fueron emitidos.

Adicionalmente, el modelo considera una cantidad variable de metadatos para cada autor, canal y manifestación, lo que facilita agregar nuevos canales a la aplicación.

El diagrama de clases del modelo completo se muestra en la figura 3.3.

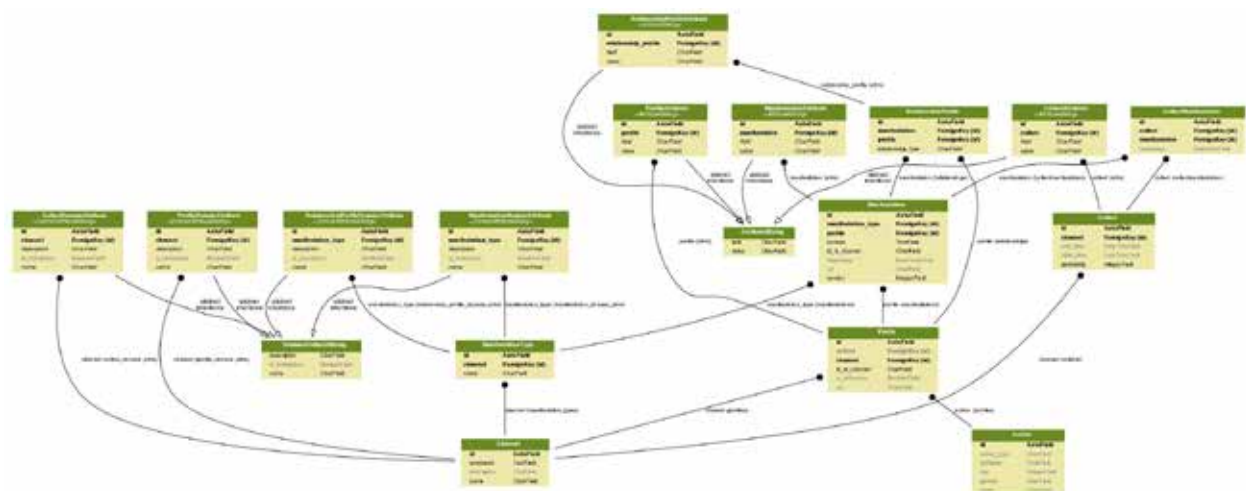


Figura 3.3: Modelo de datos de Babel

⁹ <https://babel.labhackercd.leg.br/> ⁸<https://www.djangoproject.com>

3.3.2 API

Babel expone dos interfaces para programadores que hacen que Babel pueda ser un repositorio abierto y centralizado de datos de texto de figuras públicas:

- Una API¹⁰ REST¹¹ abierta, que permite a cualquier usuario leer los datos de Babel de manera estructurada en formato JSON¹²
- Una API REST cerrada, que permite a usuarios debidamente autenticados cargar sus propios datos a la plataforma.

3.3.3 Carga de datos y análisis

Babel incluye además algoritmos que recolectan datos de discursos parlamentarios de Brasil, los que son cargados a la base de datos.

Usando la API, estos datos son leídos por una aplicación web y mostrados como una nube de palabras.

3.4 E-Democracia

E-Democracia¹³ es la plataforma de *crowdlaw* de LabHacker, que busca abrir el proceso legislativo a la ciudadanía. Cuenta con diversos subproyectos para la fomentar la transparencia y la participación popular.

Además de tener el código de todas estas herramientas en GitHub, el sitio web de E-Democracia contiene instructivos, papers y casos de uso para el despliegue de estas herramientas por otras organizaciones.

3.4.1 Audiências Interativas

Audiências Interativas¹⁴ es un servicio de vídeos de sesiones de la cámara. Permite a los usuarios participar de las sesiones en tiempo real a través de un *chat*, donde pueden hacer comentarios y preguntas que después son leídas en vivo y respondidas por los parlamentarios.

La herramienta ha tenido un gran recibimiento por la ciudadanía, realizando más de 400 audiencias virtuales durante el 2017.

3.4.2 Wikilegis

Wikilegis¹⁵ es la herramienta más usada de E-Democracia por la ciudadanía. Permite a los usuarios participar activamente del proyecto de escritura de proyectos de ley haciendo modificaciones al texto.

Los usuarios pueden agregar, editar y eliminar artículos, además de opinar sobre las modificaciones a otros usuarios. Así, los parlamentarios pueden incorporar las modificaciones más votadas en cada proyecto, a partir de un reporte generado automáticamente¹⁶.

Patrocínio de bancos públicos a clubes de futebol
PL 9622/2018

Condiciona o recebimento de patrocínio de bancos públicos pelos clubes de futebol e associações esportivas à adoção de medidas de proteção de crianças e adolescentes contra a violência sexual.

Art. 1º Esta lei acrescenta dispositivo à Lei nº 8.069, de 13 de julho de 1990, que "dispõe sobre o Estatuto da Criança e do Adolescente e dá outras providências", condicionando o patrocínio de bancos públicos a times de futebol e outras associações esportivas à assinatura de compromisso de adoção de medidas para a proteção de crianças e adolescentes, de modo a mantê-los protegidos de abusos e todas as formas de violência sexual.
12 2
EDIÇÃO (3) + ADIÇÃO (1) EXCLUSÃO (0) COMENTÁRIOS (2)

Art. 2º A Lei nº 8.069, de 13 de julho de 1990, passa a vigorar acrescida do seguinte art. 18-C:
5 0
EDIÇÃO (0) + ADIÇÃO (0) EXCLUSÃO (3) COMENTÁRIOS (0)

"Art. 18-C Clubes de futebol e outras associações esportivas só poderão receber patrocínios ou qualquer tipo de verba de bancos públicos mediante assinatura de compromisso de adoção de medidas para a proteção de crianças e adolescentes contra abusos e todas as formas de violência sexual, que deverá conter os seguintes deveres:
14 0
EDIÇÃO (1) + ADIÇÃO (1) EXCLUSÃO (1) COMENTÁRIOS (1)

I - apoio a campanhas educativas, no seu âmbito, alertando para os riscos da exploração sexual e do trabalho infantil;
15 0
EDIÇÃO (1) + ADIÇÃO (0) EXCLUSÃO (0) COMENTÁRIOS (1)

Figura 3.4: Proyecto de ley en Wikilegis

¹⁰ <https://babel.labhackerod.leg.br/api/v1>

¹¹ https://es.wikipedia.org/wiki/Transferencia_de_Estado_Representacional

¹² <https://www.json.org>

¹³ <http://www.edemocracia.leg.br>

¹⁴ <https://edemocracia.camara.leg.br/audiencias>

¹⁵ <https://edemocracia.camara.leg.br/wikilegis>

¹⁶ <https://edemocracia.camara.leg.br/wikilegis/bill/81/report>

Art. 1º Esta lei acrescenta dispositivo à Lei nº 8.069, de 13 de julho de 1990, que "dispõe sobre o Estatuto da Criança e do Adolescente e dá outras providências", condicionando o patrocínio de bancos públicos a times de futebol e outras associações esportivas à assinatura de compromisso de adoção de medidas para a proteção de crianças e adolescentes, de modo a mantê-los protegidos de abusos e todas as formas de violência sexual.

EDIÇÃO + ADIÇÃO EXCLUSÃO

Rafael Braga

Esta ~~Fica lei proibido~~ acrescenta a dispositivo concessão à Lei nº 8.069, de 13 patrocínios de por julho de 1990, que "dispõe sobre o Estatuto da Criança e do Adolescente e dá outras providências"; condicionando o patrocínio de bancos públicos a times de futebol e outras associações esportivas à assinatura de compromisso de adoção de medidas para a proteção de crianças e adolescentes; de modo a mantê-los protegidos de abusos e todas as formas de violência sexual ~~profissionais~~.

👍 0 🗨️ 2 🗨️ COMENTÁRIOS (0)

Leo Pessoa

Esta lei acrescenta dispositivo à Lei nº 8.069, de 13 de julho de 1990, que "dispõe sobre o Estatuto da Criança e do Adolescente e dá outras providências", condicionando o patrocínio de ~~bancos instituições públicas~~ públicas a times de futebol e outras associações esportivas à assinatura de compromisso de adoção de medidas para a proteção de crianças e adolescentes, de modo a mantê-los protegidos de abusos e todas as formas de violência sexual.

👍 0 🗨️ 0 🗨️ COMENTÁRIOS (0)

Figura 3.5: Ediciones de un proyecto de ley en Wikilegis

3.4.3 Expressão

Expressão¹⁷ es un foro de participación ciudadana. Los usuarios pueden conversar con otros usuarios y con parlamentarios sobre los asuntos que se discuten en el parlamento o que afectan sus vidas.

El foro cuenta con más de 400 temas de discusión activos en 24 categorías, entre las que se encuentran política, educación, salud, trabajo, derechos humanos y otras.

¹⁷ <https://edemocracia.camara.leg.br/expressao>

4. Actividades realizadas

A continuación se describen las actividades realizadas durante la pasantía de investigación entre el lunes 11 y el viernes 15 de junio, trabajando en conjunto con el equipo de LabHacker.

4.1 Lunes

Tras las presentaciones respectivas, la principal actividad del día fue acordar los temas de trabajo a tratar durante la semana. Después de una conversación preliminar, se acordaron dos metas de trabajo:

- Estudiar la posibilidad de incorporar los datos de parlamentarios de Chile a Babel.
- Aplicar algoritmos de procesamiento de lenguaje natural a los datos que ya se encuentran actualmente en Babel, como una prueba de concepto.

Ese mismo día, Paulo Henrique Araújo presentó Babel, y se inició el estudio del código y modelo de datos de la aplicación, cerrando con un tour por el *Congreso Nacional*.

4.2 Martes

La mañana del segundo día de trabajo estuvo dedicado exclusivamente a analizar las dos posibles fuentes de datos: la Cámara de Diputados¹⁸, o el Senado de Chile¹⁹

La Cámara de Diputados publica las sesiones en documentos PDF poco estructurados que son difíciles de minar. Sin embargo, el Senado publica el texto de las intervenciones de cada parlamentario en su página web, en un formato suficientemente regular como para hacer *scraping*²⁰ y poder descargar los datos.

El resto del día fue dedicado al comienzo del desarrollo del *scraper* utilizando el framework de python *scrapy*²¹, que permite desarrollar *crawlers* y *scrapers* que extraen datos estructurados a partir de páginas web.

La parte más difícil del proceso fue descubrir la estructura de las URL de los distintos sitios web, para poder generar el índice de páginas a procesar. Este punto requirió varias iteraciones de intento y error para poder encontrar la combinación correcta de parámetros que incluir en cada *request*.



Figura 4.1: Paulo presentando Babel

¹⁸ www.camara.cl

¹⁹ <http://www.senado.cl>

²⁰ es.wikipedia.org/wiki/Web_scraping

²¹ <https://scrapy.org>



Figura 4.2: Estudio conjunto del sitio web del Senado de Chile

Una vez logrado esto, los pasos restantes consisten en la definición de metadatos a minar de cada sitio web. Al final, el *crawler* realiza las siguientes tareas:

1. De la página de senadores, extrae el nombre e identificador único de cada senador.
2. Para cada senador en la lista, extrae datos como su nombre, partido político, correo electrónico y otros. Luego, para cada año entre el 2004 a la fecha, obtiene la lista de intervenciones del senador de ese año, generando varias listas de intervenciones.
3. Para cada lista de intervenciones, obtiene la URL de cada intervención.
4. De cada intervención, extrae el texto del discurso, el identificador del senador que lo está dando y la fecha.

El resultado de la operación entrega la lista de senadores y el texto de cada una de sus intervenciones de manera estructurada. El código desarrollado durante la pasantía está disponible en el GitHub de LabHacker²²

4.3 Miércoles

Habiendo minado exitosamente el texto de las intervenciones en cámara de los senadores de Chile, la mañana del miércoles fue dedicada a integrar estos datos a Babel. Esto requirió la creación de algunos objetos en la base de datos para representar el canal, pero fue realizado sin mayores problemas.

La tarde estuvo destinada a la realización de dos charlas. Primero se presentó al GobLab UAI, sus proyectos de ciencia de datos y su programa de formación de funcionarios públicos, y luego se realizó una charla sobre técnicas de *machine learning* y *Big Data*. Ambas charlas incluyeron sesiones de discusión sobre posibles aplicaciones de estas tecnologías a proyectos de LabHacker.

4.4 Jueves

El jueves fue dedicado a evaluar las técnicas de análisis de texto usadas en Babel, y a proponer nuevos análisis basados en técnicas de procesamiento de lenguaje natural.

El único análisis que hace Babel en este momento es *bag of words*. Es una técnica que consiste en eliminar palabras vacías²³ (artículos, preposiciones, pronombres, etcétera), y luego encontrar las palabras más frecuentes a través de todos los documentos.

El análisis propuesto consiste en el uso de LDA²⁴ (por *Latent Dirichlet Allocation*). En general, el algoritmo consiste en explicar conjuntos de observaciones por grupos no observados. En análisis de texto, la técnica asume que la presen-

²² https://github.com/labhackercd/babel_scripts/tree/master/discursos_chile

²³ es.wikipedia.org/wiki/Palabra_vac%C3%ADa

²⁴ https://es.wikipedia.org/wiki/Latent_Dirichlet_Allocation

cia de ciertas palabras en un documento es consecuencia de las temáticas de las que trata. En este caso, las palabras en un documento serían lo observado, y las temáticas subyacentes lo no observado.

Durante la tarde se experimentó aplicando LDA a la base de datos de Babel. Se utilizaron las librerías de python nltk²⁵ para análisis de lenguaje natural, gensim²⁶ para la construcción del modelo LDA, y pyLDavis²⁷ para visualizar el resultado. La salida es un grupo de potenciales temáticas subyacentes, cada una de las cuales está caracterizada por una cierta frecuencia de palabras.

La figura 4.4 muestra el resultado de nuestro análisis. El uso de pyLDavis permite visualizar los datos en un *dashboard* interactivo para estudiar las palabras características de cada temática. A partir de esto no es trivial determinar



Figura 4.3: Presentaciones del GobLab UAI

qué es lo característico de cada temática, pero la experimentación preliminar muestra que esta técnica es promisoria para análisis útiles, como el descubrimiento automático de tópicos de discusión, o el análisis de sentimiento.

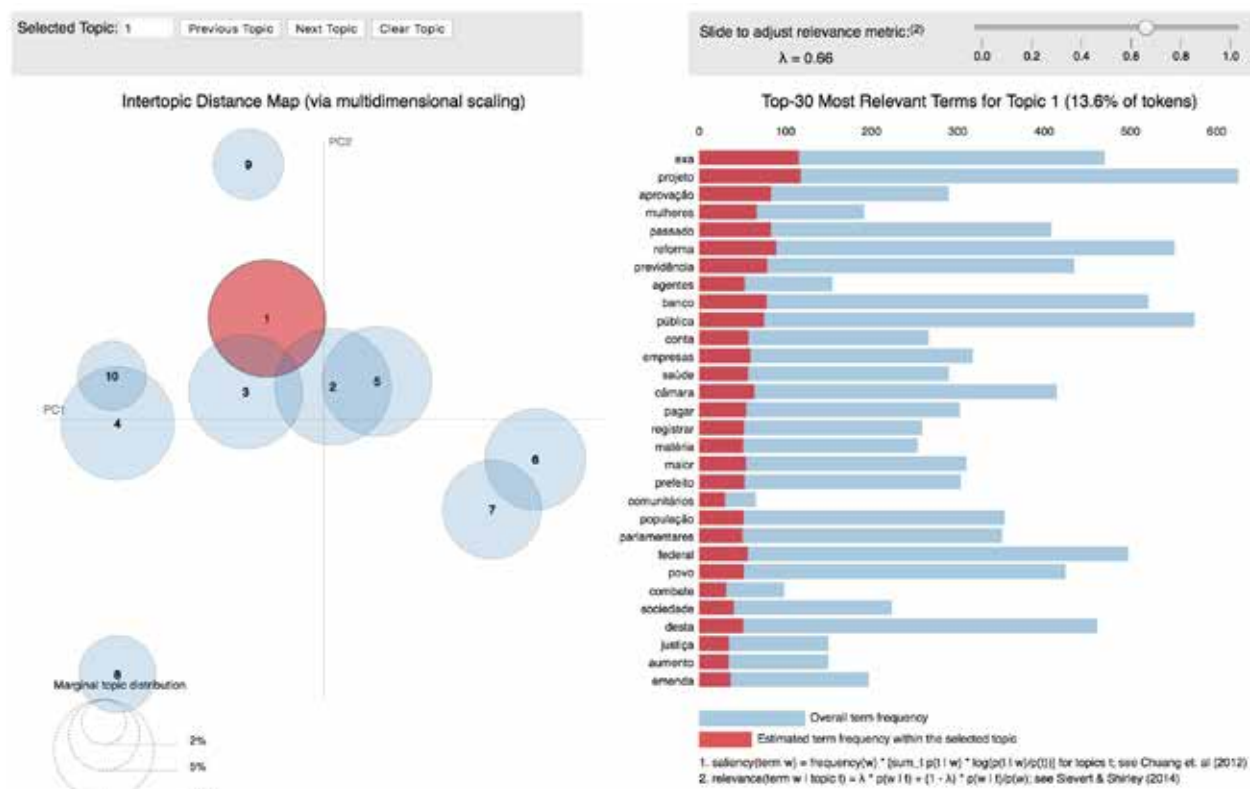


Figura 4.4: Resultado de aplicar LDA al corpus con 10 tópicos

25 <https://www.nltk.org>

26 <https://radimrehurek.com/gensim>

27 <https://github.com/bmabey/pyLDavis>

4.5 Viernes

El último día de la pasantía estuvo dedicado a conocer los otros proyectos de LabHacker, que están resumidos en el capítulo 3, y a discutir potenciales mejoras.

Wikilegis tiene un problema propio de todas las plataformas de *crowdlaw*: no todos los usuarios son igual de confiables ni igual de expertos. Para esto, se propuso como posible solución el uso de PageRank²⁸, el algoritmo de Google para asignar puntajes de importancia relativa a entidades en una red. En el algoritmo, cada entidad vota por aquellas otras entidades que cree que son confiables, con lo que se genera un peso para cada una. En cada iteración siguiente, cada entidad vuelve a votar, pero esta vez con su propio peso.

En este caso, los usuarios podrían votar unos por otros sobre cuán confiables creen que los otros son. La idea de fondo es que la opinión de alguien confiable pesa más que la de alguien que no lo es. Entonces, si alguien con alto peso considera que un usuario es confiable, es más probable que efectivamente lo sea.

Dada la falta de tiempo la idea no pudo ser implementada durante la pasantía, pero parece ser una propuesta interesante.

²⁸ <https://en.wikipedia.org/wiki/PageRank>

5. Conclusiones

Los objetivos específicos se lograron con creces: se buscaba principalmente identificar la factibilidad de incorporar datos del Senado chileno a Babel. En lugar de esto, se minaron los datos y se integraron a la aplicación exitosamente.

En consecuencia, para usar Babel en Chile bastaría con traducir la aplicación al castellano, cosa que en django es fácil de hacer. Sin embargo, probablemente sea más útil internacionalizar la instancia de Babel de LabHacker en vez de lanzar uno paralelo: la API permite cargar datos de manera distribuida, y mantener centralizado el repositorio de datos tiene beneficios para los investigadores que lo deseen usar. Más aún, el tener los datos en un solo repositorio y en un mismo esquema facilitaría reproducir los mismos análisis para datos de Brasil y Chile, lo que potencialmente incentivaría a investigadores de otros países a sumar sus datos.

Wikilegis también puede ser utilizado en Chile internacionalizando el código sin mayores dificultades. Esto ignora el problema de efectivamente conseguir que las aplicaciones sean utilizadas por los actores relevantes, que sobre todo en el caso de Wikilegis es particularmente difícil, pero al menos la barrera técnica es superable.

Por otra parte, nuestra experimentación muestra que el uso de técnicas de *machine learning* y procesamiento de lenguaje natural permite hacer análisis interesantes sobre los datos de Babel, pero requieren una dedicación importante por parte del equipo de desarrollo.

Sobre los problemas de confiabilidad en los usuarios asociados a *crowd law*, el uso de técnicas de análisis de redes sociales parece tener potencial para mitigarlos o eliminarlos, pero requiere más estudio y experimentación.

A nivel más general, ambos centros se beneficiaron con este intercambio, conociendo los proyectos del otro y creando instancias de colaboración futura. Por esto, la principal conclusión de la pasantía es que la colaboración entre laboratorios de investigación de gobierno es altamente productiva, sobre todo cuando sus áreas de especialización son complementarias.

Editan:



P o l i L a b



 **Lab Hacker**

