Percepciones sobre el mercado laboral en Ñuble a través de Facebook

Una aplicación sencilla de minería de texto

Héctor Garrido Henríquez - Observatorio Laboral Ñuble

11 de agosto de 2017

Motivación

- Más allá de las estadísticas oficiales, las personas generan mucha información que está disponible en la Web
- El uso de redes sociales, en particular Facebook, ha fomentado las interacciones sociales entre desconocidos de manera explosiva
- ► Es común que las personas interactúen en grupos asociativos para vender artículos, arrendar viviendas y **buscar empleo**

No todo lo que brilla es oro

- A pesar de que las redes sociales generan grandes volúmenes de información en tiempo real (Big Data). No reemplazan las estadísticas oficiales de empleo porque
- No existe forma (razonable) de verificar que los posts sean fidedignos
- Un requerimiento de empleo puede postearse muchas veces (duplicidad)
- Las redes sociales no representan a toda una población pues hay aún ciertos grupos que no las utilizan. De forma que de ninguna forma reemplazan una encuesta representativa

Lo anterior implica que utilicemos las redes sociales para medir percepciones, más que cualquier otra cosa

Contexto

- ▶ Para extrear los posts sobre empleo es necesario que los grupos sean: (i) públicos o (ii) ser administrador de dicho grupo.
- Los grupos en Facebook son por defecto cerrados. De los grupos chillanenses con actividad importante todos son cerrados.
- Me contacte con un usuario y le pedí que me cediera la administración de un grupo, explicándole mi labor en el OLR Ñuble. Dicha persona accedió amablemente.
- Esto ahorra mucho trabajo en la medida en que el grupo ya tenía actividad y muchos miembros.

Datos de trabajo en Chillán

- Dirección del grupo: https://www.facebook.com/groups/241293439569652/
- ▶ 17.542 miembros
- 4.974 posts en un período aproximado de un año.
- ▶ El grupo no tiene posts (o casi) sobre ventas u otras temáticas, pues el administrador anterior fue muy meticuloso en mantener el objetivo del grupo, labor que he seguido realizando, revisando éste a diario y borrando aquellos posts fuera de contexto

Metodología

- Facebook provee una AF través de la cual se pueden descargar posts, ya sea de grupos o páginas.
- A través del lenguaje R utilicé la librería Rfacebook para conectarme con dicha API
- Puedo descargar además de los posts (como texto, directamente), las fechas y las reacciones a dichos posts (likes, reacciones, shares, etc)

Metodología

- ▶ Una vez obtenidos los posteos es necesario depurar y darle cierta estructura al texto.
- Básicamente se eliminan números, símbolos, puntuaciones y emojis.
- Parte importante de la preparación de los datos consiste en eliminar stopwords, que son palabras inútiles para nuestro objetivo.
- ► También se reemplazan sinónimos.

A continuación se presentan algunos resultados preliminares

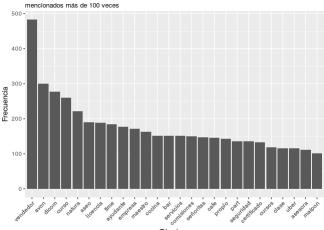
Word Cloud



realizan senora 🤗 e adultocel currículum bio informacion interesa cebollino basevia conductor, promotoria interesa cebollino promotoria interesa pollo diarios g cositaalto verifica_{fox} saludos promotora instalación pre operador operador hablar hija construcción serena chofer atento garzona g queso construcción serena chote chote chote construcción serena chote ampliaciones básica urgencia varon psi viajes cumplir estare saben frey Cursos comision carmen lame cajaventa clase empresa cuidar auxiliar S sushi porfavor luis internet demanda 2 quillota santiago ect co terreno tura cocina cuidat a fuero vehículos toda IICencia cafe o w sin cuidando Servicios sino pega ta o cortés bulnes guardias concepción consumos tormal viaje grupo mayores wifi ángeles deuda bono buenos tentacion mocopia manejo puedo fundación enegocio kilo técnico viejo realizados dar maestra agotas de maestra agotas ademas canales kit frambuesas cosechado tondo tres sueños empieza conocimientos info marzo calidad gasfiteria valdivia mesera

Los términos más frecuentes (i)

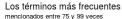


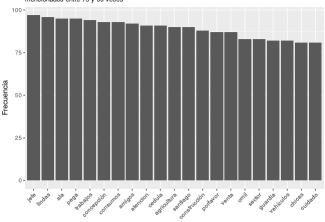


Términos

Elaboración propia a partir de datos de Facebook

Los términos más frecuentes (ii)

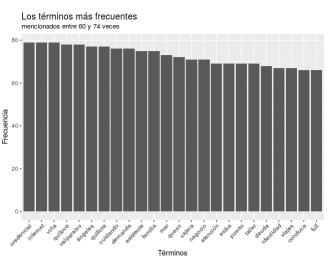




Términos

Elaboración propia a partir de datos de Facebook

Los términos más frecuentes (iii)



Elaboración propia a partir de datos de Facebook

Asociación entre palabras

- En minería de texto las palabras pueden asociarse entre sí, de manera similar a lo que entendemos como correlación en estadística.
- ► Entenderemos que dos conceptos están asociados en un texto en la medida en que sea frecuente que aparezcan juntos.
- Un ejemplo bastante obvio de esto podrían ser los términos "part" y "time"

Curso



La palabra curso está asociada a los siguientes términos

cedula	varon	varones	credencial	omil
0.73	0.73	0.73	0.73	0.73
fotoc	certificado	afp	copia	afilicacion
0.73	0.72	0.72	0.72	0.72
confeccion	manipulador	recaudadora	tasador	dama
0.72	0.72	0.72	0.72	0.71
identidad	fonasa	certif	finiquito	afiliacion
0.70	0.70	0.70	0.69	0.68
minima	reflejada	damas	alimento	idealmente
0.68	0.68	0.67	0.63	0.61
electromecanico	agronomo	antigüedad	barmant	boletin
0.61	0.61	0.61	0.61	0.61
brigadistas	certifocado	cocineros	conduccion	corresponde
0.61	0.61	0.61	0.61	0.61
currcilum	empleador	idelamente	monotores	plc
0.61	0.61	0.61	0.61	0.61
sonidista	telefonoy			
0.61	0.61			

Licencia

0.60

La palabra licencia está asociada a los siguientes términos

Tonasa	атр	Tiniquito	TOTOC	conductor	certificado	cedula
0.74	0.73	0.72	0.72	0.71	0.70	0.70
varon	varones	hoja	credencial	omil	agronomo	antigüedad
0.70	0.70	0.70	0.70	0.70	0.70	0.70
barmant	boletin	brigadistas	certifocado	cocineros	conduccion	corresponde
0.70	0.70	0.70	0.70	0.70	0.70	0.70
currcilum	empleador	idelamente	monotores	plc	sonidista	telefonoy
0.70	0.70	0.70	0.70	0.70	0.70	0.70
antigua	dama	afilicacion	confeccion	manipulador	recaudadora	tasador
0.68	0.67	0.67	0.67	0.67	0.67	0.67
clase	copia	identidad	afiliacion	certif	alimento	minima
0.66	0.66	0.65	0.63	0.63	0.62	0.61
idealmente						

Lo que viene

A través de reglas de asociación y análisis de frecuencias, y la propia interpretación de ciertos textos, se pretende entregar información sobre qué ocupaciones generan más discusión en Ñuble

- Si bien, el grupo es el caballito de batalla, se recurrirá a otras fuentes como las fan pages de las OMIL de la provincia
- ► También se extraerá información de otros grupos.