



UCSC



The Fifteenth International Conference on Information, Process, and Knowledge Management - eKNOW 2023

Tools Based on Word Embedding to Make Easy the Analysis
of Emotions in Spanish Text

Jorge Silva Pedreros, Alejandra Segura-Navarrete, Christian Vidal-Castro
jorge.silva1602@alumnos.ubiobio.cl, asegura@ubiobio.cl, cvidal@ubiobio.cl
Universidad del BíoBío, Chile

Claudia Martínez-Araneda
cmartinez@ucsc.cl
Universidad Católica de la Santísima Concepción, UCSC, Chile



TOPICS

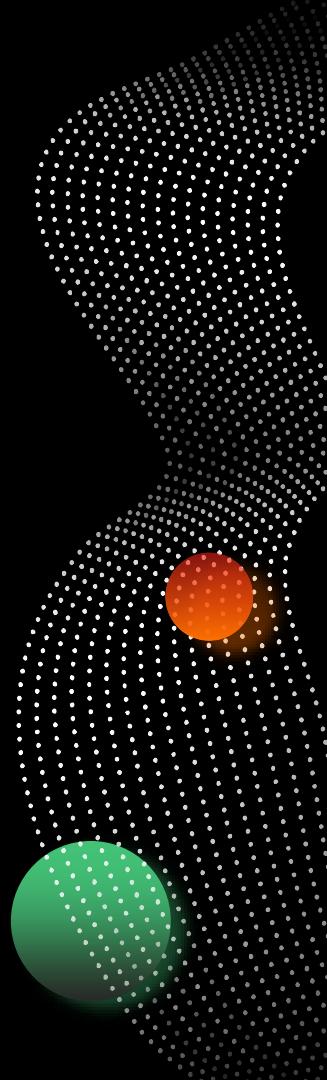
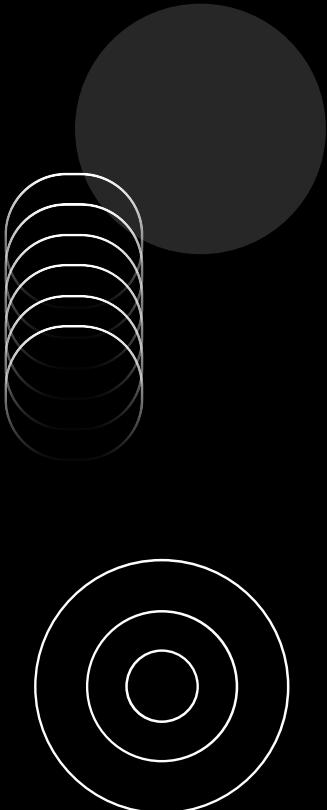
01 Analysis of emotions

02 Background (Tools, WE, fastText)

03 Machine Lerning process

04 Integrate tool development

05 Conclusions and future work

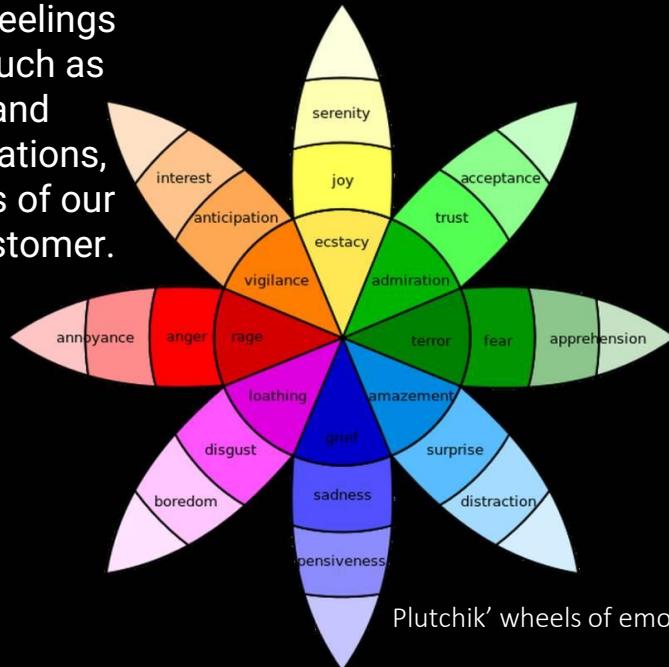


Analysis of emotion

01



Def. Emotion Analysis is a process that aims to detect and recognize types of feelings through the expression of texts, such as anger, disgust, fear, joy, sadness and surprise. It can have useful applications, such as: measuring the happiness of our citizens or the experience of a customer.

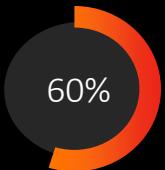


01

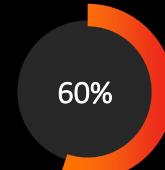
Background - Tools

- Synesketch
- EMTk
- Receptiviti
- IBM Watson NL
- Understanding Text Analysis
- Text2Emotion
- ...among others

Preprocessing integrated



Open source



Analysis approach

Lexicon or dictionary



Machine Learning

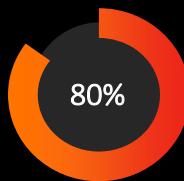


SVM
NB
RF

Hybrid

10%

English



Language supported

English, French

10%

Spanish

10%

02 Background – We and fasttext

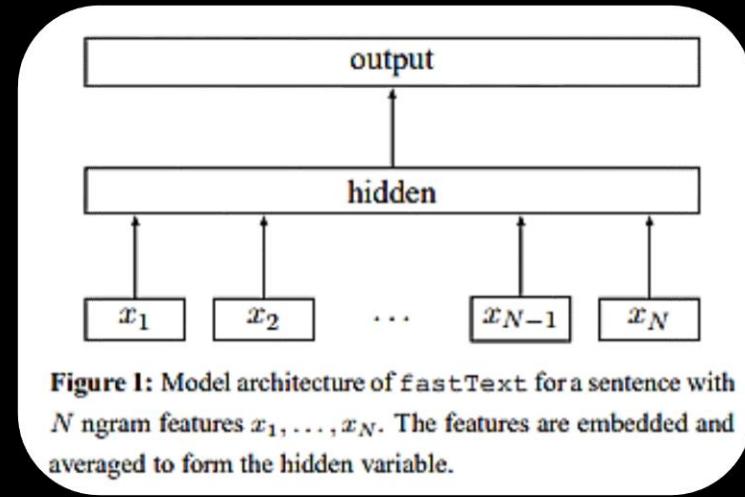
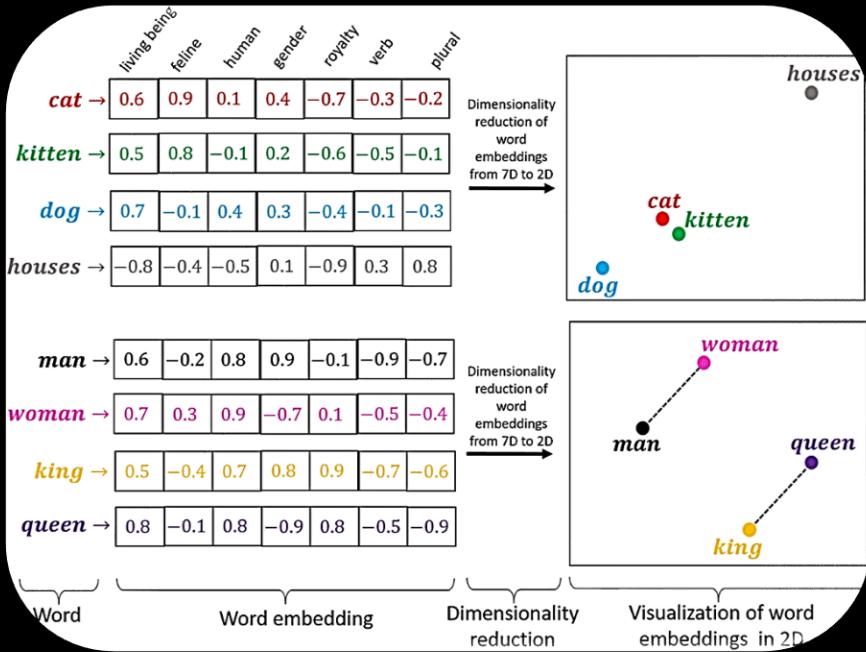


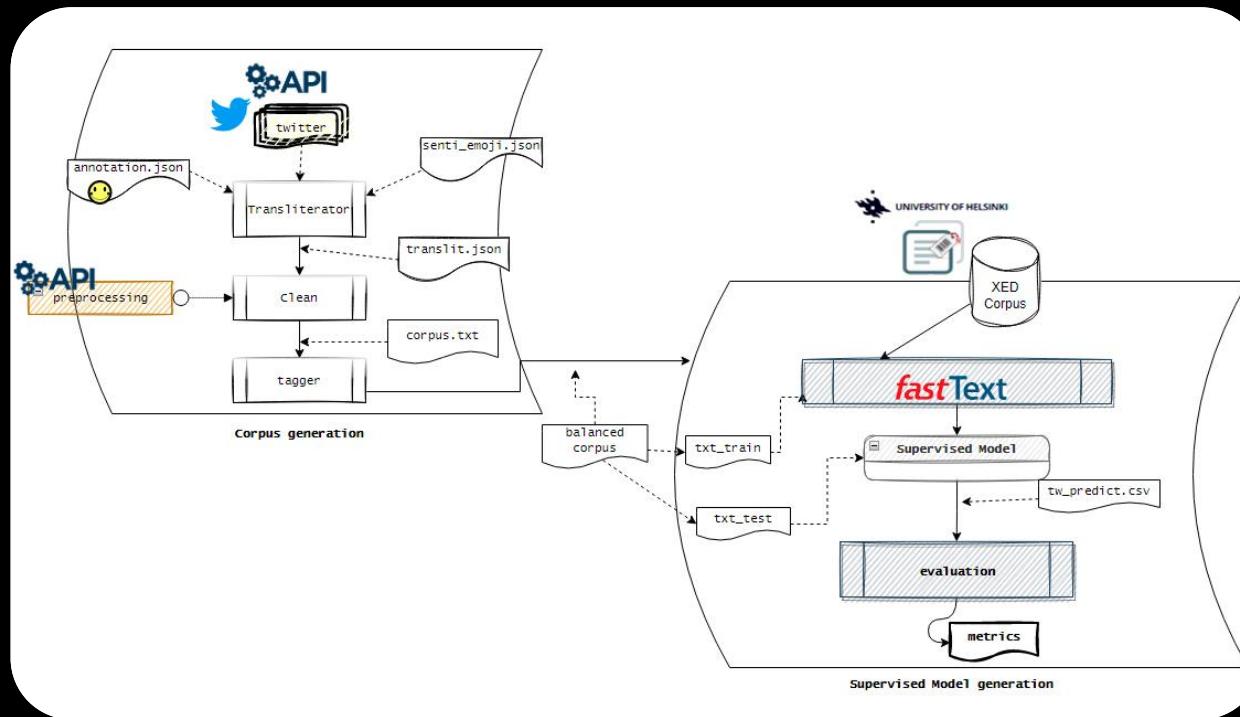
Figure 1: Model architecture of fastText for a sentence with N n-gram features x_1, \dots, x_N . The features are embedded and averaged to form the hidden variable.

Joulin, Grave, Bojanowski, & Mikolov, T. (2016).

$$model^*(ngrams, learningrate, epochs) = P(model, ngrams, learningrate, epochs, trainingdata)$$

03

Machine learning process



XED available from
<https://github.com/Helsinki-NLP/XED>

03

Preliminary results

Joy			Surprise			fear			anger			sadness			disgust			anticipation			trust		
P@	R@	f1	P@	R@	f1	P@	R@	f1	P@	R@	f1	P@	R@	f1	P@	R@	f1	P@	R@	f1	P@	R@	f1
0.34	0.46	0.27	0.26	0.33	0.21	0.16	0.21	0.12	0.45	0.43	0.46	0.27	0.33	0.22	0.19	0.32	0.13	0.37	0.43	0.33	0.25	0.39	0.19

P@: Precision

R@: Recall

F1: F-measure

04

Integrated tool – fastText embedding



cuac.dev
Tesis

A code has been emailed to you.

Enter code

Sign in

← Back

Resend email

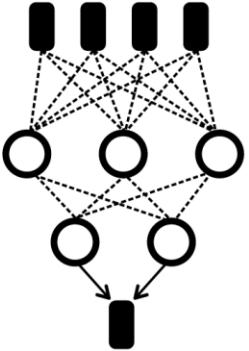
Didn't receive an email?

Cloudflare Zero Trust

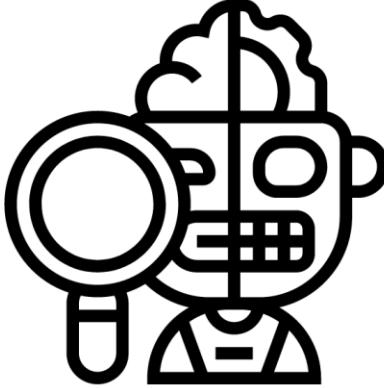
Tesis Word Embeddings Preprocesamiento Word Embedding Análisis de Emociones



Preprocesamiento
Limpieza de corpus para ML o NLP.



Word Embedding
Obtén palabras similares o resuelve analogías en español.



Análisis de Emociones
Detecta emociones de la rueda de emociones de Plutchik en oraciones en español.

Spanish version

04 Integrated tool – preprocessing module

Preprocesamiento de Corpus

Herramienta para el preprocesamiento de texto de desarrollo propio, código fuente y herramienta CLI disponible en [Repositorio Tesis](#) (privado)

Corpus

No se ha seleccionado ningún archivo

Formatos de archivo soportados: texto plano o HTML

Tamaño máximo: ~100 MiB

Opciones

Limpiar etiquetas HTML

Selector CSS para retirar etiquetas

script, style, pre, svg, math, noscript, ref

Selector CSS para concatenar separador de línea

p, hr, br, li, tr

Quitar acentos, caracteres de control y convertir letras mayúsculas en minúsculas

Quita signos de puntuación y reemplazar tokens extraños (dígitos o URLs) con <unk/>

04

Integrated tool – similarity example

```
$ ./fasttext similarity result/fil9.bin #command line
```

```
Pre-computing word vectors... done.
```

Query word? Asparagus

beetroot 0.812384

tomato 0.806688

horseradish 0.805928

spinach 0.801483

licorice 0.791697

lingonberries 0.781507

asparagales 0.780756

lingonberry 0.778534

celery 0.774529

beets 0.773984

Integrated tool – similarity & analogy module

Word Embeddings Preprocessing Word Embedding Analogy Embedding

Word Embedding

Word embedding de FastText (skip-gram) de

Similaridad

Palabra

incendios

Ver similares

Parámetros Utilizados para E

- minn: 3
- maxn: 8
- lr: 0.01

Resultados

Palabra	Similaridad
forestales	0.913796
siniestros	0.831326
incendio	0.718994
forestal	0.706075
conaf	0.705643
llamas	0.697386
focos	0.694914
antiincendios	0.691427
pastizales	0.68594
vegetacion	0.679587

Cerrar

The screenshot shows a mobile application interface. At the top, there are tabs for 'Word Embeddings', 'Preprocessing', 'Word Embedding', and 'Analogy Embedding'. Below this, the title 'Word Embedding' is displayed, followed by a subtitle 'Word embedding de FastText (skip-gram) de'. A section titled 'Similaridad' contains a text input field with the word 'incendios' and a blue button labeled 'Ver similares'. Below this, a section titled 'Parámetros Utilizados para E' lists parameters: 'minn: 3', 'maxn: 8', and 'lr: 0.01'. A large modal window titled 'Resultados' is open, listing words and their similarity scores. The list includes: forestales (0.913796), siniestros (0.831326), incendio (0.718994), forestal (0.706075), conaf (0.705643), llamas (0.697386), focos (0.694914), antiincendios (0.691427), pastizales (0.68594), and vegetacion (0.679587). A close button 'X' is at the top right of the modal, and a 'Cerrar' button is at the bottom right.

04

04

Integrated tool – analogy example

```
$ ./fasttext analogies result/fil9.bin #command line
```

Pre-computing word vectors... done.

Query triplet (A - B + C)? berlin germany france

paris 0.896462

bourges 0.768954

louveciennes 0.765569

toulouse 0.761916

valenciennes 0.760251

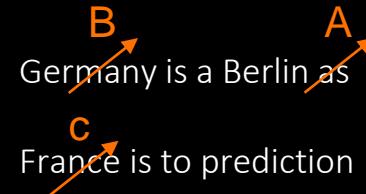
montpellier 0.752747

strasbourg 0.744487

meudon 0.74143

bordeaux 0.740635

pigneaux 0.736122



Integrated tool – similarity & analogy module

Word Embedding

Word embedding de FastText (skip-gram) de desarrollo propio a partir de artículos de prensa chilena.

Resultados	
Palabra	Similaridad
dilma	0.681452
rousseff	0.653195
mandataria	0.631464
michelle	0.630353
lula	0.616748
roussef	0.598940
inacio	0.594999
luiz	0.587153
rousseff	0.584675

Parámetros

- minn:
- maxn:
- lr: 0.0:
- minCo

Analogía	
A	bachelet
B	chile
C	brasil

Ver analogías

- ws: 23
- neg: 11
- thread: 4
- dim: 300

B is to A as C is to **prediction**

Chile is to Bachelet as Brasil is to **prediction**

Chile is to Bachelet as Brasil is to **Dilma**

04

Integrated tool – analysis of motions module

Análisis de Emociones

Modelo FastText supervisado con parámetros idénticos al [Word Embedding](#) (salvo WordNgrams = 17 y loss=ns) a partir de Tweets con emojis y hashtags particulares.

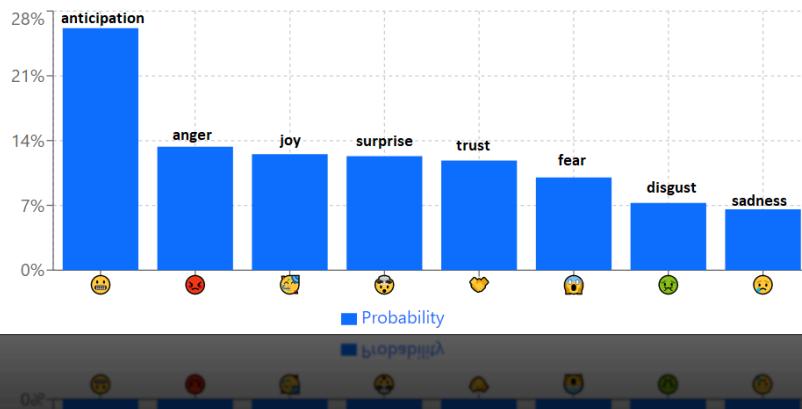
Nótese que pueden ser ingresadas múltiples oraciones y cada una tendrá su análisis individual, por lo que puede, por ejemplo, pegar un artículo y ver el análisis emotivo de cada oración.

Texto a analizar

```
##@@@Mañana iré a conocer una tienda IKEA !!! :) :)
```

Analizar Texto

manana ire a conocer una tienda ikea



04

04

Evaluation of integrated tool



	Preprocessing	S & A	Analysis of Emotion
Performance	93%	96%	96%
Accessibility	87% 	89% 	90% 
Best Practices	100% 	100% 	100% 
SEO	92%	100% 	100% 

by Google lighthouse

05

Conclusions and future work

- Both the API for text preprocessing and the web application for analysis of emotions based on fastText are available in GitHub repository.
- It would be of interest to create a web browser extension to provide feedback on the predominant emotions in the text that a user writes or to hide comments or news articles in which negative emotions predominate.
- Another aspect that could be incorporated into future work would be how a recommendation system could be nourished from the results of the similarity module and analogies such as the analysis of emotions.

REFERENCES

- M. Lepe-Faúndez, A. Segura Navarrete, C. Vidal-Castro, C. Martínez-Araneda, and C. Rubio-Manzano, "Detecting aggressiveness in tweets: A hybrid model for detecting cyberbullying in the Spanish language," *Applied Sciences*, 11(22), 10706, 2021.
- A. Joulin, E. Grave, P. Bojanowski, and T. Mikolov, "Bag of tricks for efficient text classification," *arXiv preprint arXiv:1607.01759*, 2016.
- P. Bojanowski, P. Grave, A. Joulin, and T. Mikolov, "Enriching word vectors with subword information," *Transactions of the Association for Computational Linguistics*, 5, pp. 135–146, 2017, doi: https://doi.org/10.1162/tacl_a_00051
- M. Bekoff. "Animal Emotions: Exploring Passionate NaturesCurrent interdisciplinary research provides compelling evidence that many animals experience such emotions as joy, fear, love, despair, and grief—we are not alone". *BioScience* 50(10), pp. 861–870, 2000.
- R. C. Solomon, and B. Duignan. Emotion [Encyclopedia]. Encyclopædia Britannica, Available from <https://www.britannica.com/science/emotion>, last accessed March 2023.
- A. Yadollahi, A. G. Shahraki, and O. R. Zaiane, "Current state of text sentiment analysis from opinion to emotion mining," *ACM Computing Surveys (CSUR)* 50(2), pp. 1–33, 2017.
- P.J. Stone, D.C. Dunphy, and M. S. Smith. "The general inquirer: A computer approach to content analysis," M.I.T Press.
- S. Mohammad, and P. Turney. "Crowdsourcing a word–emotion association lexicon". *Computational intelligence* 29(3), pp. 436–465, 2013.
- R. Plutchik. "A general psychoevolutionary theory of emotion". In *Theories of emotion*, pp. 3–33. Academic Press, 1980.
- R. Varela. Affective Lexical Resource to Spanish Language (Recurso Léxico Afectivo para idioma Español: REDPAL). Master Thesis. Universidad del Bío-Bío. [Online]. Available from http://mcc.ubiobio.cl/docs/tesis/roberto_varela_medina-2020_varela_roberto.pdf, last accessed March 2023.
- E. Öhman, M. Pàmies, K. Kajava, and J. Tiedemann, "XED: A Multilingual Dataset for Sentiment Analysis and Emotion Detection," In *Proceedings of the 28th International Conference on Computational Linguistics*, pp. 6542–6552, 2020. <https://doi.org/https://doi.org/10.18653/v1/2020.coling-main.575>
- S. Wang, A. Maoliniyazi, X. Wu, and X. Meng, "Emo2Vec: Learning emotional embeddings via multi-emotion category," *ACM Transactions on Internet Technology (TOIT)* 20(2), pp. 1–17, 2020.
- E. Cambria, A. Livingstone, and A. Hussain, "The hourglass of emotions," In *Cognitive behavioural systems*, pp. 144–157. Springer, Berlin, Heidelberg, 2012, doi: https://doi.org/10.1007/978-3-642-34584-5_11

REFERENCES

- P. Gupta, R. Inika, B. Gunnika, and A. K. Dubey, "Decoding emotions in text Using GloVe embeddings," In International Conference on Computing, Communication, and Intelligent Systems (ICCCIS), pp. 36–40. IEEE, 2021, doi: <https://doi.org/10.1109/ICCCIS51004.2021.9397132>
- B. Saju, J. Siji, and A. Amal, "Comprehensive Study on Sentiment Analysis: Types, Approaches, Recent Applications, Tools and APIs," In Advanced Computing and Communication Technologies for High Performance Applications (ACCTHPA 2020), pp. 186–193. IEEE, 2020.
- U. Krcadinac , P. Pasquier, J. Jovanovic, and V. Devedzic, "Synesketch: An open-source library for sentence-based emotion recognition," IEEE Transactions on Affective Computing 4(3), pp. 312–325, 2013.
- Receptiviti, Inc. Receptiviti's Emotions engine. [Online]. Available from <https://docs.receptiviti.com/frameworks/emotions>, last accessed March 2023.
- F. Calefato, F. Lanobile, and N. Novielli, "EmoTxt: A toolkit for emotion recognition from text," In 2017 seventh International Conference on Affective Computing and Intelligent Interaction Workshops and Demos (ACIIW), pp. 79–80. IEEE, 2017.
- F. Calefato, F. Lanobile, N. Novielli, and L. Quaranta, "EMTk: The Emotion Mining Toolkit," SEmotion'19, pp. 34–37, 2019, doi: <https://dl.acm.org/doi/10.1109/SEmotion.2019.00014>
- IBM. IBM Watson Natural Language Understanding Text Analysis. [Online]. Available from <https://www.ibm.com/demos/live/natural-language-understanding/self-service/home>, last accessed December 2022.
- A. Band. Text2emotion: Python package to detect emotions from textual data. [Online]. Available from <https://pypi.org/project/text2emotion/>, last accessed March 2023.
- J. Silva. Corpus-preproc v0.1.0: A preprocessor for text and HTML corpora. [Online]. Available from <https://crates.io/crates/corpus-preproc>, last accessed March 2023.
- J. Silva, A. Segura Navarrete. FastText-embedding-viewer: An emotion detect tool based on fastText [Online]. Available from <http://tesis.cuac.dev>, last accessed March 2023.
- A.M.J. Skulimowski, "Cognitive Content Recommendation in Digital Knowledge Repositories – A Survey of Recent Trends," In: L. Rutkowski, M. Korytkowski, R. Scherer, R. Tadeusiewicz, L. Zadeh, J. Zurada (eds) Artificial Intelligence and Soft Computing. ICAISC 2017. Lecture Notes in Computer Science, vol 10246. Springer, Cham, 2017. https://doi.org/10.1007/978-3-319-59060-8_52

Team



Christian Vidal-Castro

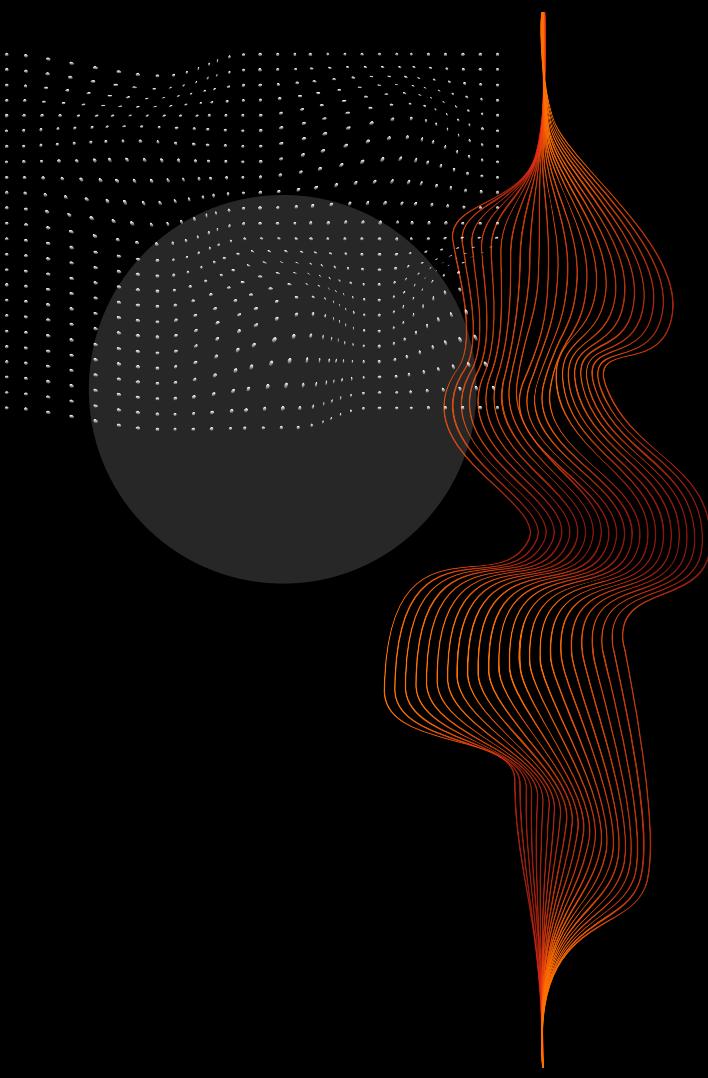
chvidal@ubiobio.cl

Alejandra Segura Navarrete

asegura@ubiobio.cl

Jorge Silva Pedreros

jorge.silva1602@alumnos.ubiobio.cl



Thanks!

any questions?