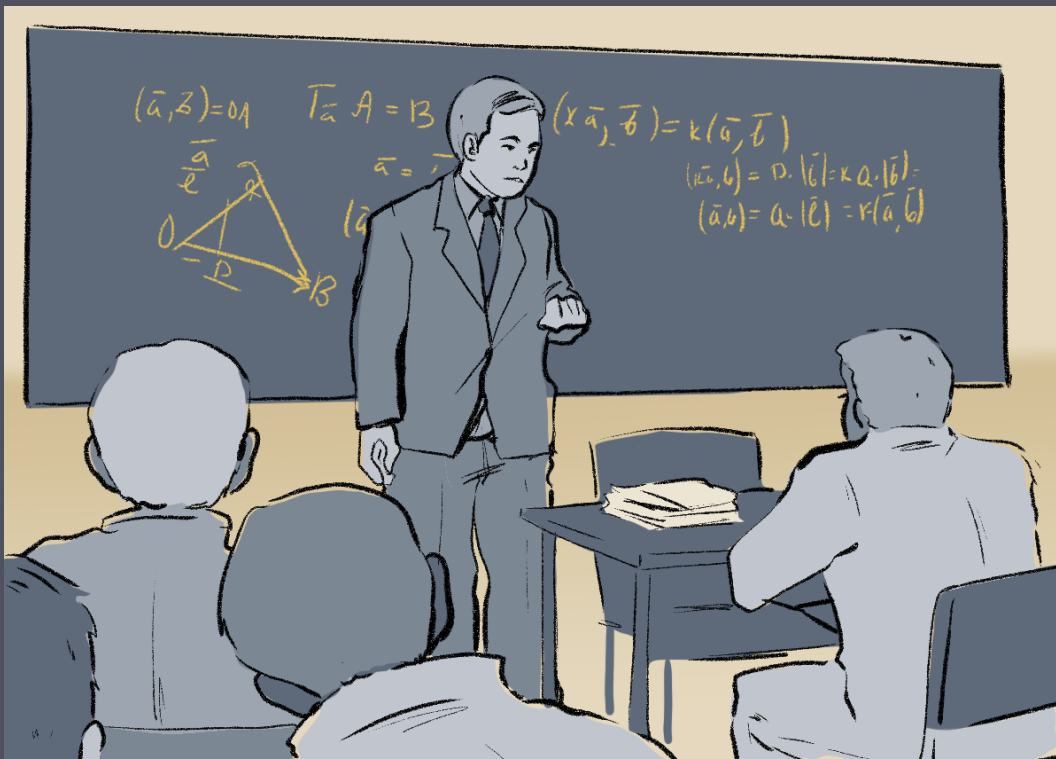


# BOURBAKI

COLEGIO DE MATEMÁTICAS

Procesamiento del Lenguaje Natural  
en 25 artículos





Esta colección de 25 artículos académicos ha sido seleccionada para acompañar el curso de **Procesamiento de Lenguaje Natural (NLP)**, orientado al estudio intermedio de modelos lingüísticos y técnicas matemáticas aplicadas a datos textuales. El curso recorre desde representaciones básicas como bolsa de palabras y Tf-idf, hasta arquitecturas avanzadas como Transformers, BERT, GPT y estrategias de ajuste fino en grandes modelos del lenguaje.

A lo largo del temario se abordan módulos clave: análisis de sentimientos, técnicas de topic modeling (incluyendo LDA), redes densas y embeddings como Word2vec y Doc2vec, redes recurrentes y modelos secuenciales (LSTM, Seq2Seq y atención de Bahdanau), y finalmente la exploración de grandes modelos del lenguaje con aplicaciones contemporáneas como RAG y destilación.

Los artículos incluidos en esta bibliografía combinan fundamentos teóricos con aplicaciones prácticas, proporcionando al estudiante una visión profunda de cómo el aprendizaje automático se articula con la lingüística computacional. Se ha priorizado literatura que ilustre tanto la evolución de las representaciones vectoriales del lenguaje como el impacto de los modelos generativos en tareas de clasificación, resumen, traducción y personalización.

Esta selección busca situar los contenidos del curso dentro del marco de la investigación moderna en NLP, conectando cada módulo con hitos relevantes en el desarrollo de arquitecturas y metodologías. Al mismo tiempo, ofrece a los participantes referencias clave para aplicar estas técnicas en sectores como la salud, la educación, los servicios financieros y el diseño de productos inteligentes, promoviendo una visión crítica, ética y aplicada de la inteligencia artificial

# 1.

---

Shannon, C.E. (1948) 'A mathematical theory of communication', Bell System Technical Journal, 27, pp. 379–423 & 623–656.

# 2.

---

Harris, Z.S. (1954) 'Distributional structure', Word, 10(2–3), pp. 146–162.

# 3.

---

Luhn, H.P. (1958) 'The automatic creation of literature abstracts', IBM Journal of Research and Development, 2(2), pp. 159–165.

# 4.

---

Spärck Jones, K. (1972) 'A statistical interpretation of term specificity and its application in retrieval', Journal of Documentation, 28(1), pp. 11–21.

# 5.

---

Salton, G. and Buckley, C. (1988) 'Term-weighting approaches in automatic text retrieval', Information Processing & Management, 24(5), pp. 513–523.

# 6.

---

Robertson, S.E. and Walker, S. (1994) 'Some simple effective approximations to the 2-Poisson model for probabilistic weighted retrieval', In: Proceedings of SIGIR '94, pp. 232–241.

# 7.

---

Deerwester, S., Dumais, S.T., Furnas, G.W., Landauer, T.K. and Harshman, R. (1990) 'Indexing by latent semantic analysis', Journal of the American Society for Information Science, 41(6), pp. 391–407.

# 8.

---

Church, K.W. (1988) 'A stochastic parts program and noun phrase parser for unrestricted text', In: Proceedings of the 2nd Conference on Applied Natural Language Processing, pp. 136–

# 9.

---

Brown, P.F., Della Pietra, S.A., Della Pietra, V.J. and Mercer, R.L. (1993) 'The mathematics of statistical machine translation: Parameter estimation', Computational Linguistics, 19(2), pp. 263–311.



# 10.

---

Papineni, K., Roukos, S., Ward, T. and Zhu, W.-J. (2002) 'BLEU: a method for automatic evaluation of machine translation', In: Proceedings of ACL 2002, pp. 311–318.

# 11.

---

Lin, C.-Y. (2004) 'ROUGE: A package for automatic evaluation of summaries', In: Proceedings of the Workshop on Text Summarization Branches Out, ACL 2004.

# 12.

---

Kneser, R. and Ney, H. (1995) 'Improved backing-off for m-gram language modeling', In: Proceedings of ICASSP 1995, vol. 1, pp. 181–184.

# 13.

---

Bengio, Y., Ducharme, R., Vincent, P. and Jauvin, C. (2003) 'A neural probabilistic language model', Journal of Machine Learning Research, 3, pp. 1137–1155.

# 14.

---

Hochreiter, S. and Schmidhuber, J. (1997) 'Long short-term memory', Neural Computation, 9(8), pp. 1735–1780.

# 15.

---

Pang, B., Lee, L. and Vaithyanathan, S. (2002) 'Thumbs up? Sentiment classification using machine learning techniques', In: Proceedings of EMNLP 2002, pp. 79–86.

# 16.

---

Mikolov, T., Sutskever, I., Chen, K., Corrado, G. and Dean, J. (2013) 'Distributed representations of words and phrases and their compositionality', In: Advances in Neural Information Processing Systems 26 (NeurIPS 2013), pp.

# 17.

---

Pennington, J., Socher, R. and Manning, C.D. (2014) 'GloVe: Global vectors for word representation', In: Proceedings of EMNLP 2014, pp. 1532–1543.

# 18.

---

Le, Q. and Mikolov, T. (2014) 'Distributed representations of sentences and documents', In: Proceedings of ICML 2014, pp. 1188–1196.



# 19.

---

Blei, D.M., Ng, A.Y. and Jordan, M.I. (2003) 'Latent Dirichlet Allocation', *Journal of Machine Learning Research*, 3, pp. 993–1022.

# 20.

---

Bahdanau, D., Cho, K. and Bengio, Y. (2015) 'Neural machine translation by jointly learning to align and translate', In: *Proceedings of ICLR 2015*.

# 21.

---

Sutskever, I., Vinyals, O. and Le, Q.V. (2014) 'Sequence to sequence learning with neural networks', In: *Advances in Neural Information Processing Systems 27 (NeurIPS 2014)*, pp. 3104–3112.

# 22.

---

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł. and Polosukhin, I. (2017) 'Attention is all you need', In: *Advances in Neural Information Processing Systems 30 (NeurIPS 2017)*, pp. 5998–6008.

# 23.

---

Devlin, J., Chang, M.-W., Lee, K. and Toutanova, K. (2019) 'BERT: Pre-training of deep bidirectional transformers for language understanding', In: *Proceedings of NAACL-HLT 2019*, pp. 4171–4186.

# 24.

---


Peters, M.E., Neumann, M., Iyyer, M., Gardner, M., Clark, C., Lee, K. and Zettlemoyer, L. (2018) 'Deep contextualized word representations', In: *Proceedings of NAACL-HLT 2018*, pp. 2227–2237.

# 25.

---

Brown, T.B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., et al. (2020) 'Language models are few-shot learners', In: *Advances in Neural Information Processing Systems 33 (NeurIPS 2020)*, pp. 1877–1901.





Estaremos encantados de explicar  
con detalle sobre nuestros cursos

Escríbenos por mail o WhatsApp

[info@colegio-bourbaki.com](mailto:info@colegio-bourbaki.com)

+52 56 2141 7850

[colegio-bourbaki.com](http://colegio-bourbaki.com)

**B O U R B A K I**

COLEGIO DE MATEMÁTICAS