# Natural Language Processing (NLP)

Coimbra, April 2024

Nikitas N. Karanikolas
Dept. of Informatics and Computer Engineering
University of West Attica, Athens, Greece
nnk@uniwa.gr
http://users.uniwa.gr/nnk

# Natural Language Processing (NLP)

- Summarization
- Text Classification
- Macine Translation
- Question Answering
- Text Generation
- Computer Assisted Assessment
- Sentiment Analysis
- Opinion Mining
- Subjectivity Analysis
- Corpus Building
- Named Entity Recognition
- Chatbots – Virtual Assistants
- Image Captioning

# Summarization [1, 2]

- Can be Comprehensive (semantic oriented)  or Extractive (shallow processing)
- Extractive is based on the selection of the most prominent sentences to convey the meaning. It is based on:
  - Weight of words (TF-IDF, TF-ISF, TF-RIDF)
  - Sentence Location (Baxentale, News Articles)
  - Title Words
- See also my previous lecture "Extractive summarization", June 2017, Belgrade
https://users.uniwa.gr/nnk/InternationalLectures/Extractive%20summarization%20June%202017.pdf

# Text Summarization Applications

➢ **News Organizations:** Automatic generation of brief news article summaries for quick comprehension.

➢ **Researchers:** Condensing academic papers for rapid identification of key findings.

➢ **Content Aggregation Platforms and Websites:** Creating digests aiding user article selection based on interests.

➢ **Finance Sector:** Summarizing financial reports for swift evaluation by investors and analysts.

➢ **Healthcare Professionals:** Summarizing medical conversations, records, and research papers for streamlined tasks.

➢ **Social Media Insights:** Summarizing discussions and reviews for understanding public sentiment and reactions.

➢ **Market Research:** Distilling consumer feedback and reviews for prompt trend identification.

➢ **Educational Support:** Creating concise educational material summaries for student comprehension.

# Text Classification [3]

- Assign to a documents a class label (the category that the document belongs)

- Can be based on the existence of words or phrases

- The method needs training and training data (pre-classified documents)

- It is Critical to create an Authority List of words or phrases that will be appropriate to discriminate between classes

- See also:
https://users.uniwa.gr/nnk/InternationalLectures/Text%20classification%20based%20on%20phrases%20June%202017.pdf

# Text Classification Applications

➢ **Email Sorting:** Distinguishing spam from genuine emails by evaluating content and sender details.

➢ **E-commerce Taxonomy:** Organizing products and multi-level taxonomy trees.

➢ **Social Media for Sentiment Analysis:** Assessing user-generated content for sentiments - positive, negative, or neutral.

➢ **Personalized Recommendations:** Tailored suggestions in e-commerce platforms.

➢ **Healthcare Data:** Classifying medical records and patient data to enhance organization and support research.

➢ **Legal Document Classification:** Simplifying document management for legal firms by classifying contracts and case files.

➢ **Fraud Detection in Finance:** Identifying suspicious transactions and user behavior for fraud detection.

# Machine Translation Applications [10]

- **Translation Services:** Text, documents, websites, and spoken language.

- **Content Creation and Marketing:** Translating content into multiple languages.

- **Government and Diplomatic Use:** Translating official documents, treaties, and diplomatic communications.

- **News Agencies:** Swiftly translating news articles to provide timely and accurate information in multiple languages.

- **E-commerce and Retail:** Translating product descriptions, reviews, and checkout processes.

- **International Business Communications:** Translating contracts, emails, and documentation for negotiations and collaborations.

- **Travel Industry Integration:** Providing multilingual information for tourists through apps and websites.

- **Customer Support:** Facilitating communication between international customers and support teams.

- **Language Learning Platforms:** Offering language courses and translation assistance to enhance language skills.

# Question Answering [11] – Applications

- **Company Chatbots:** Addressing customer inquiries, providing information, and troubleshooting.
- **Voice-Activated Assistants:** assist users and execute tasks.
- **E-commerce:** Enabling proactive information seeking for informed purchasing decisions.
- **E-Learning Platforms:** Addressing student queries, providing explanations and solutions.
- **Healthcare:** Assisting healthcare professionals with clinical queries, symptoms, and generating reports.
- **Legal Sector:** Providing insights into case law, statutes, and regulations for legal research.
- **Technical Support:** Offering guidance and solutions for software and hardware queries.

# Text Generation [12] – Applications

- **Revolutionizing Industries:** Automating processes across diverse sectors.
- **Content Creation:** product descriptions, ads, promotions, news articles and blogs.
- **Financial Reporting Automation:** Analyzing market data to automatically generate financial reports.
- **Educational Contribution:**
  - Facilitating e-learning by producing quizzes, study materials, and explanations.
  - Enhancing engagement and personalized learning experiences.
- **Creative Writing Support:** Contributing to narratives, short stories, and poetry.
- **Chatbot & Virtual Assistant Roles:**
  - Empowering chatbots with natural language responses for customer inquiries.
  - Assisting users in tasks, providing information, and troubleshooting issues.

# Computer Assisted Assessment [4 - 6]

- Mechanically assign a grate to an answer with respect to the expected (correct) answer.

- There is a need for positive training (correct answers and textbook) data and negative training data (erroneous answers)

- Can be based on phrases

# Sentiment Analysis, Opinion Mining, Subjectivity Analysis [add Ref.]

- Subjectivity Analysis
classify a given text as subjective or objective

- Sentiment Analysis or Polarity Analysis
Once a text is subjective
Assign a score Positive or Negative

- Affective Computing
Attempt to identify emotional charge
  - Happiness
  - Sadness
  - Fear
  - Anger - Passion

# Sentiment Analysis Applications

➢ **Social Media and Customer Feedback:** Assessing sentiments to manage reputation and enhance product quality.

➢ **Diverse Data Sources:** Processing surveys, forms, and reviews to gauge customer satisfaction.

➢ **Financial Sector:** Analyzing news, reports, and social media for market sentiment.

➢ **Polling and Political Forecasting:** Gauging public sentiment for informed decision-making in politics.

➢ **Entertainment Industry:** Evaluating audience reactions to movies and TV shows.

➢ **Product Launches:** Assessing public sentiment to evaluate market reception.

➢ **Healthcare:** Understanding patient sentiment to drive improvements in patient care.

➢ **Social Media:** Measuring campaign effectiveness and tracking brand trends.

➢ **News Content Tailoring:** Assessing reader reactions to tailor preferences.

➢ **Customer Support Insights:** Identifying dissatisfied customers.

➢ **Travel Services:** Analyzing reviews to enhance services and increase bookings.

# Corpus Building [7, 8]

- Dialectal lexicon building
  See my previous lecture "Dialectal lexicon building: requirements and technical specifications", Belgrade June 2017
  https://users.uniwa.gr/nnk/InternationalLectures/Dialectal%20lexicon%20building%20June%202017.pdf

- Dialectal corpora building
  See my previous lecture "Dialectal Corpora Building (for oral and written sources)", Belgrade June 2017
  https://users.uniwa.gr/nnk/InternationalLectures/Dialectal%20Corpora%20Building%20June%202017.pdf

# Named Entity Recognition [13]

- to be expanded

- See the presentaion by Cvetana Krstev,
  Department of Library and Information
  Sciences, Faculty of Philology, University of
  Belgrade, Serbia
  in the following link:
  https://users.uniwa.gr/nnk/InternationalLectures/NamedEntities.pdf

# Chatbots [9] & Virtual Assistants Applications

➢ **E-commerce & Online Services:** Chatbots can assist customers, provide product details, and resolve issues instantly.

➢ **Banking & Finance:** Virtual assistants aid with inquiries and financial guidance.

➢ **Travel:** Chatbots facilitate bookings, offer travel information.

➢ **Healthcare:** Virtual assistants provide medical information and schedule appointments.

➢ **Language Learning:** Virtual tutors assist in practicing languages.

➢ **Human Resources:** Chatbots aid with HR-related inquiries and company policies.

➢ **Online Retail:** Chatbots offer personalized product suggestions.

➢ **Business Reception:** Virtual receptionists manage calls and provide information.

➢ **Education:** Chatbots can assist with course-related queries and assignments.

➢ **News Updates:** Chatbots deliver tailored news content and updates to users.

➢ **Smart Home Devices:** Voice-activated assistants are used for inquiries, smart device control, and information provision.

# Image Captioning Applications [add Ref.]

➢ **Social Media Platforms:** Use NLP to create image captions, benefiting user engagement and accessibility, particularly for the visually impaired.

➢ **E-commerce:** Image captioning in catalogs aids product management and improves search experiences.

➢ **Education:** Educational platforms use image captions to explain visual content, assisting students in understanding complex visuals.

➢ **Healthcare:** Image captioning assists in describing medical images, aiding in diagnostics and conveying findings among medical professionals.

➢ **Image Search and Retrieval:** Image captioning in search engines helps users find images by describing them in text queries.

➢ **Content Moderation:** Social media platforms use image captioning to identify inappropriate or harmful content, ensuring a safer online environment.

➢ **News Content:** News agencies utilize image captioning to enhance the accessibility of news articles for readers.

# References

- [1] **Nikitas N. Karanikolas**, Eleni Galiotou and Christodoulos Tsoulloftas, A workbench for extractive summarizing methods. PCI'2012: 16th Panhellenic Conference on Informatics, October 5-7, 2012, Piraeus, Greece. IEEE CPS. https://ieeexplore.ieee.org/document/6377346/

- [2] **Nikitas N. Karanikolas**, "Extractive summarization methods – subtitles and method combinations". RTA-CSIT 2016, November 18 - 19, 2016, Tirana, Albania. http://ceur-ws.org/Vol-1746/paper-06.pdf

- [3] **N. Karanikolas** and C. Skourlas. A parametric methodology for text classification. Journal of Information Science, Vol. 36 (4), pp. 421-442, 2010, doi:10.1177/0165551510368620. http://journals.sagepub.com/doi/10.1177/0165551510368620

- [4] **Nikitas N. Karanikolas**. The role of phrases in Information Retrieval and related domains. eRA-4. 4th Conference for the contribution of Information Technology to Science, Economy, Society and Education, September 25-26, 2009, Spetses, Greece.
http://users.teiath.gr/nnk/papers/B22_cr.pdf

- [5] **Nikitas N. Karanikolas**, Computer Assisted Assessment (CAA) of Free-Text: Literature Review and the specification of an alternative CAA system. In 19th IEEE International Workshops on Enabling Technologies: Infrastructures for Collaborative Enterprises (WETICE 2010), June 28-30, 2010, Larissa, Greece, IEEE Xplore
https://ieeexplore.ieee.org/document/5541991/

- [6] **Nikitas N. Karanikolas**, "Summarization as the base for Text Assessment". 4th IC-ININFO, September 5-8, 2014, Madrid, Spain.
https://aip.scitation.org/doi/abs/10.1063/1.4907844

- [7] **Nikitas N. Karanikolas**, Eleni Galiotou, George J. Xydopoulos, Angela Ralli, Konstantinos Athanasakos, and George Koronakis, Structuring a Multimedia Tri-dialectal Dictionary. TSD'2013: Text, Speech and Dialogue, 16th International Conference, September 1-5, 2013, Pilsen, Czech Republic. Springer, LNAI 8082, DOI: 10.1007/978-3-642-40585-3_64.

- [8] **Nikitas N. Karanikolas**, Eleni Galiotou, Dimitris Papazachariou, Konstantinos Athanasakos, George Koronakis and Angela Ralli, "Towards a computational processing of oral dialectal data". PCI 2015, October 01 - 03, 2015, Athens, Greece. ACM 978-1-4503-3551-5, DOI: 10.1145/2801948.2801966.

- [9] Eirini Manga, **Nikitas Karanikolas**, Catherine Marinagi and Christos Skourlas. Evaluating Citizen Comments in Public Consultations Using Data Mining: Analyzing Legislation Comments for the Greek General Commercial Registry. PCI 2021: 430-435. https://dl.acm.org/doi/10.1145/3503823.3503902

- [10] Vanlalmuansangi Khenglawt, Lalṭanpuia. *Machine translation and its approaches*. Mizoram Science Congress 2018 (MSC 2018). Advances in Engineering Research, vol. 178, December 2018. DOI: 10.2991/msc-18.2018.22. https://www.atlantis-press.com/article/55909743.pdf

- [11] Abdelghani Bouziane, Djelloul Bouchiha, Noureddine Doumi, Mimoun Malki. *Question Answering Systems: Survey and Trends*. Advanced Wireless, Information, and Communication Technologies (AWICT 2015). Procedia Computer Science, vol. 73, 2015, Pages 366-375. DOI: 10.1016/j.procs.2015.12.005. https://www.sciencedirect.com/science/article/pii/S1877050915034663

- [12] Ehud Reiter. 1996. *Building Natural-Language Generation Systems*. Arxiv: cmp-lg/9605002. https://arxiv.org/pdf/cmp-lg/9605002

- [13] Wikipedia. *Named-entity recognition*. https://en.wikipedia.org/wiki/Named-entity_recognition