PART 1

• For each period, assign a name to each generated topic based on the topic's top words. **List all topic names** in your report. If a topic is incoherent to the degree that no common theme is detectable, you can just mark it as incoherent (i.e., no need to name a topic that does not exist).

To control the word count in each file, we modified the download process to set a maximum limit of 50,000 words. We tried three different approaches to preprocessing those datasets.

For Titles Before 1990: We remained with the basic preprocessing method, which involved converting all characters to lowercase and removing unnecessary numbers and symbols.

For Titles From 1990 to 2009: In this period, we implemented language detection for each title. Upon identifying the language, **titles from different languages were translated into English**. This step was followed by the standard preprocessing routine of tokenization, stop word removal, lowercasing, lemmatization, numerical characters removal.

For Titles From 2010 Onwards: For this recent period, we continued with language detection for each title. Based on the detected language, we used a language-specific stopword removal approach. After this, the preprocessing steps include character lowercasing and normalization, using *'unicodedata.normalize()'* function in Python, ensuring uniformity and compatibility in Unicode representations, here we used this function for characters(different languages) that might have different forms but represent the same text.

Topics Summarization:

Before 1990
5 topics:
Topic 1: Systems Design and Control
Topic 2: Analysis of Parallel Algorithms and Models
Topic 3: Computer Logic and Programming
Topic 4: Problem and Application on machine and network
Topic 5: Information Processing and Structures

From 1990 to 2009
5 topics:
Topic 1:Information Modeling and Network Systems
Topic 2:Performance Analysis of Digital Systems
Topic 3:Control and Design using CMOS and Fuzzy Logic
Topic 4:Study of Voltage and Logic Speed
Topic 5:Power Management and Decision Support Systems

10 topics:
Topic 1:Neural Networks and Sensor Evaluation
Topic 2:Control and Performance Analysis of Networks
Topic 3:Method and Architecture for Problem Solving
Topic 4:High-Speed Memory and Video Quality

Topic 5:Data Management and Image Techniques
Topic 6:Information Systems and Real-Time Support
Topic 7:Fuzzy Approach and Dynamic Modeling
Topic 8:Systems Learning and Signal Processing
Topic 9:CMOS Design and Web Development
Topic 10:Power and Digital Circuit Simulation

From 2010
5 topics
Topic 1:Systems engineering and control
Topic 2:Computational analysis and algorithms
Topic 3:Computer science and theoretical computing
Topic 4:Algorithmic methods and applications
Topic 5:Information systems and graph theory

10 topics
Topic 1:Adaptive network design and mobile technologies
Topic 2:Wireless control and cooperative algorithms
Topic 3:Field knowledge and optimal computation
Topic 4:Information systems and hybrid vehicle optimization
Topic 5:Communication systems and MIMO technologies
Topic 6:Machine learning and online gaming
Topic 7:Data analysis and dynamic computing
Topic 8:Energy modeling and web services
Topic 9:Performance modeling and simulation
Topic 10:Resource allocation and state analysis

• Do the topics make sense to you? Are they coherent? Do you observe trends? **Discuss in 4-6 sentences**.

The titles across these three groups seem to be coherent to some degree. For titles before 1990, It seems that the topics are related to the field in the foundation of computer science, such as Systems design and information structures, which is appropriate when concerning the time. For titles from 1990 to 2009, they were more concentrating on network or model, subsequently, for titles from 2010 onwards, the topics show a clear evolution towards more sophisticated and integrated aspects of computer science. In summary, the topics transition from foundational concepts to a more complex and data-centric approach in the most recent decade.

PART2

• Again: Assign a name to each topic based on the topic's top words (for each period). **List all topic names in your report.**

From 1990:
Topic 1: Formal Logic and Automata
Topic 2: Environmental and Multiobjective Systems
Topic 3: Algorithmic Problem-solving and Pattern Recognition

Topic 4: Computer Science and Information Systems
Topic 5: Systems Analysis and Decision Modeling

From 1990 to 2009:
Topic 1: Information Retrieval and AI in User Behavior
Topic 2: Numerical Solutions and Programmatic Approaches
Topic 3: High-Frequency Power Conversion in CMOS
Topic 4: Fuzzy Neural Network-Based System Modeling
Topic 5: Real-Time Performance Analysis in Wireless Mobile Networks

From 2010 onwards:
Topic 1: Social Behavior Research and Evidence
Topic 2: Wireless Network Performance and Resource Allocation
Topic 3: Simulation Modeling and Optimization Methods
Topic 4: Platoon Control and Sampled Data Sequences
Topic 5: Vehicle Control and Learning Approaches Using Data Models

• Bianchi et al. 2021 claim that their approach produces more coherent topics than previous methods. **Let's test this claim by comparing the coherence of the topics produced by CTM with the topics produced by LDA.** Describe your observations in 3-4 sentences.

After comparing the two sets of results in five topics, it is evident that while both sets of topics are relevant, CTM topics are notably more specific and contextual, whereas LDA topics tend to be broader and less specific. Additionally, CTM demonstrates a higher degree of topic coherence compared to LDA, particularly in representing complex and abstract subjects. The specificity and coherence of CTM's results make them more understandable and intuitive to humans.

• Do the two models generate similar topics? Can you discover the same temporal trends (if there are any)? Discuss in 5-6 sentences.
The two models generate similar topics related to computer science, but they differ in terms of specific focus areas. For example, in the time periods before 1990, CTM appears to have a broader range of topics, including theoretical aspects (formal logic, automata) and applied areas (environmental systems, decision modeling), whereas LDA seems more focused on practical aspects such as systems design, algorithm analysis, and computer programming.

In the topics generated by LDA, we can find a shift from foundational theoretical topics in computer science and systems analysis in 1990 towards applied and specialized areas, particularly focusing on social behavior, wireless networks, simulation, and control systems from 1990 to 2009 and onwards. In the topics generated by CTM, the trend across the three time periods indicates a progression from foundational topics in systems design and computer logic before 1990, towards a focus on applied areas such as information modeling and digital systems performance analysis from 1990 to 2009, followed by a continued evolution into more advanced and diverse fields, including systems engineering, and computational analysis from 2010 onwards. Overall, the 2 models share similar temporal trends from foundational theoretical topics to advanced topics in the field of Computer Science.