

SENTENCE RECONSTRUCTION FOR THE KAGGLE SANTA 2024 COMPETITION



UNIVERSIDAD
FRANSISCO JOSE DE
CALDAS

GABRIELA MARTÍNEZ SILVA, JAIRO ARTURO BARRERA MOSQUERA

INTRODUCTION

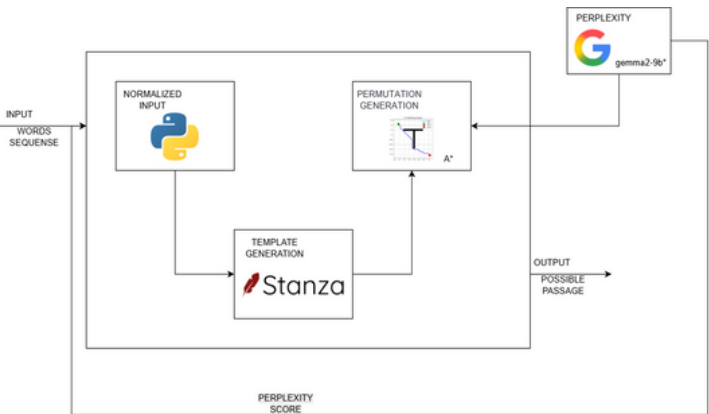
Reconstructing coherent sentences from scrambled word sequences is a fundamental challenge in Natural Language Processing (NLP). With a permutation space that grows factorially (e.g., over 3.6 million for 10 words), This work presents a modular and linguistically informed system to reorder words efficiently. It focuses on input normalization and syntactic-morphological analysis to extract grammatical constraints. These constraints guide a search algorithm that prunes the permutation space, generating only linguistically plausible arrangements, aiming to optimize resources and improve coherence, applied to the context of the "Kaggle Santa 2024" competition

METHODOLOGY

- The proposed system is structured into 3 main modules and evaluation feedback to reconstruct sentences from scrambled word lists:
1. Input Normalization:
 - Basic cleaning consistent formatting is performed,
 2. Template Definition:
 - Receives the normalized word list.
 - Applies syntactic and morphological analysis .
 - Extracts: Soft structural templates: Abstract patterns (e.g., [NOUN] [VERB] [ADJECTIVE]).
 3. Permutation Generator:
 - Uses templates and constraints to intelligently explore the permutation space, avoiding exhaustive enumeration.
 - Algorithms like A* Search (states as partial sequences) are employed.

RESULTS

The primary result of this work is the successful design and detailed architecture of a modular system for reconstructing coherent sentences from scrambled word lists, specifically conceptualized for challenges like the "Kaggle Santa 2024" competition



The implemented architecture ensures a sequential flow of data, where each module builds upon the output of the previous one. This pipeline from raw input normalization, through linguistic constraint extraction and informed permutation generation, to robust evaluation

DISCUSSION

- Strengths:
- Efficiency: Pruning the search space based on linguistic constraints significantly reduces computational load.
 - Interpretability: Grammatical constraints offer an understandable basis for the system's decisions.
 - Modularity: Allows for the substitution or improvement of individual components
- Potential Limitations:
- Dependence on the quality of NLP tools for initial analysis.
 - Scalability to extremely long sentences or highly specialized vocabularies if templates become too complex.

CONCLUSIÓN

The presented systemic design offers a promising approach for the efficient and coherent reconstruction of sentences from scrambled words. By integrating linguistic analysis for generating constraints and templates, the system minimizes computational resources and focuses on grammatically plausible permutations. It is expected that the implementation and evaluation of this system in the "Kaggle Santa 2024" competition will validate its superiority in terms of reconstruction quality and efficiency compared to less structured methods