# Form 3: Methodology

## 1.Team No: 12

**2.Project Title:** A Verifiable Semantic Searching Scheme by Optimal Matching over Encrypted Data in Public Cloud.

## 3. Proposed Method:

In this paper, we propose a secure verifiable semantic searching scheme that treats matching between queries and documents as an optimal matching task. We treat the document words as “suppliers,” the query words as “consumers,” and the semantic information as “product,” and design the minimum word transportation cost (MWTC) as the similarity metric between queries and documents. Therefore, we introduce word embeddings to represent words and compute Euclidean distance as the similarity distance between words, then formulate the word transportation (WT) problems based on the word embeddings representation. However, the cloud server could learn sensitive information in the WT problems, such as the similarity between words. For semantic optimal matching on the ciphertext, we further propose a secure transformation to transform WT problems into random linear programming (LP) problems. In this way, the cloud can leverage any readymade optimizer to solve the RLP problems and obtain the encrypted MWTC as measurements without learning sensitive information. Considering the cloud server may be dishonest to return wrong/forged search results, we explore the duality theorem of linear programming (LP) and derive a set of necessary and sufficient conditions that the intermediate data produced in the matching process must satisfy. Thus, we can verify whether the cloud solves correctly RLP problems and further confirm the correctness of search results.

## 4.Proposed Method Illustration:

This approach views documents' words as "suppliers," query words as "consumers," and semantic information as the "product," with the Minimum Word Transportation Cost (MWTC) serving as the similarity metric. Word embeddings are employed to represent words, and the Euclidean distance computes the similarity distance between them, forming the basis for subsequent computations.

To address the risk of the cloud server learning sensitive information from the Word Transportation (WT) problems, such as word similarity, the method introduces a secure transformation. This transformation converts WT problems into Random Linear Programming (RLP) problems, safeguarding sensitive semantic details. By leveraging any available optimizer, the cloud server can solve the RLP problems and obtain encrypted MWTC measurements without accessing sensitive information directly.

However, considering the potential for dishonest behavior from the cloud server, the method incorporates the duality theorem of linear programming (LP) to establish a

verification mechanism. This mechanism derives a set of necessary and sufficient conditions that intermediate data during the matching process must satisfy. By verifying these conditions, the method can confirm whether the cloud server has correctly solved the RLP problems, ensuring the correctness of the search results. Overall, this approach ensures both security and verifiability in semantic searching, safeguarding sensitive information while maintaining the integrity of search outcomes, even in less trustworthy cloud environments.

## 5.Parameters and Formulas:

1. Word Embeddings:
    - Representation of words in a high-dimensional vector space capturing semantic relationships.
    - Parameters include vector dimensions, training algorithms (e.g., Word2Vec, GloVe), and pre-trained embeddings.
2. Euclidean Distance:
    - Metric used to compute similarity between word embeddings.
    - Parameters include distance threshold for similarity determination and normalization techniques.
3. Minimum Word Transportation Cost (MWTC):
    - Metric representing the optimal matching cost between queries and documents based on semantic information.
    - Parameters include cost calculation methods and threshold values for determining similarity.
4. Word Transportation (WT) Problems:
    - Formulated based on word embeddings to optimize semantic transportation between words.
    - Parameters include optimization algorithms (e.g., linear programming), constraints, and objective functions.
5. Random Linear Programming (RLP) Problems:
    - Secure transformation of WT problems to prevent leakage of sensitive semantic information.
    - Parameters include transformation algorithms, randomness parameters, and encryption techniques.