The problem is an Entity Recognition problem. You have to use python and any relevant libraries to solve this, as long as you understand underlying assumptions and methods (at least on a high level) being used. The preferred solution would include a jupyter notebook with all your analysis(EDA on dataset is recommended) and experiments accompanying the submission file.

The data is annotated for named entities:

1. Person
2. Location
3. Group
4. Creative work
5. Corporation
6. Product

**Input:**

train.txt

test.txt

**Output:**

submission.txt

You should train your model(s) on train.txt and compute the most probable entity label on data from test.txt. Evaluation of the model will be done on the test labels (which are not provided to you). The submission file should have Entity and corresponding label separated by a tab(\t).

Please submit a well-documented jupyter notebook containing all your analysis, solution and experiments (if any) along with the submission file. Also note down all the approaches you can think of, its advantage and disadvantages/limitations.

Let me know if you have any follow up questions.

Bonus Points: If you can give confidence score of a prediction along with entity predicted it would be great. 0 being - not confident, 1 being - very confident.