



Then and Now Map

Mohammed Eltaher , Jeongkyu Lee , Shuyang Li , Hau-Wen Chang+, Dongwon Lee+
 University of Bridgeport
 The Pennsylvania State University+
<http://www.bridgeport.edu/~jelee/mia>

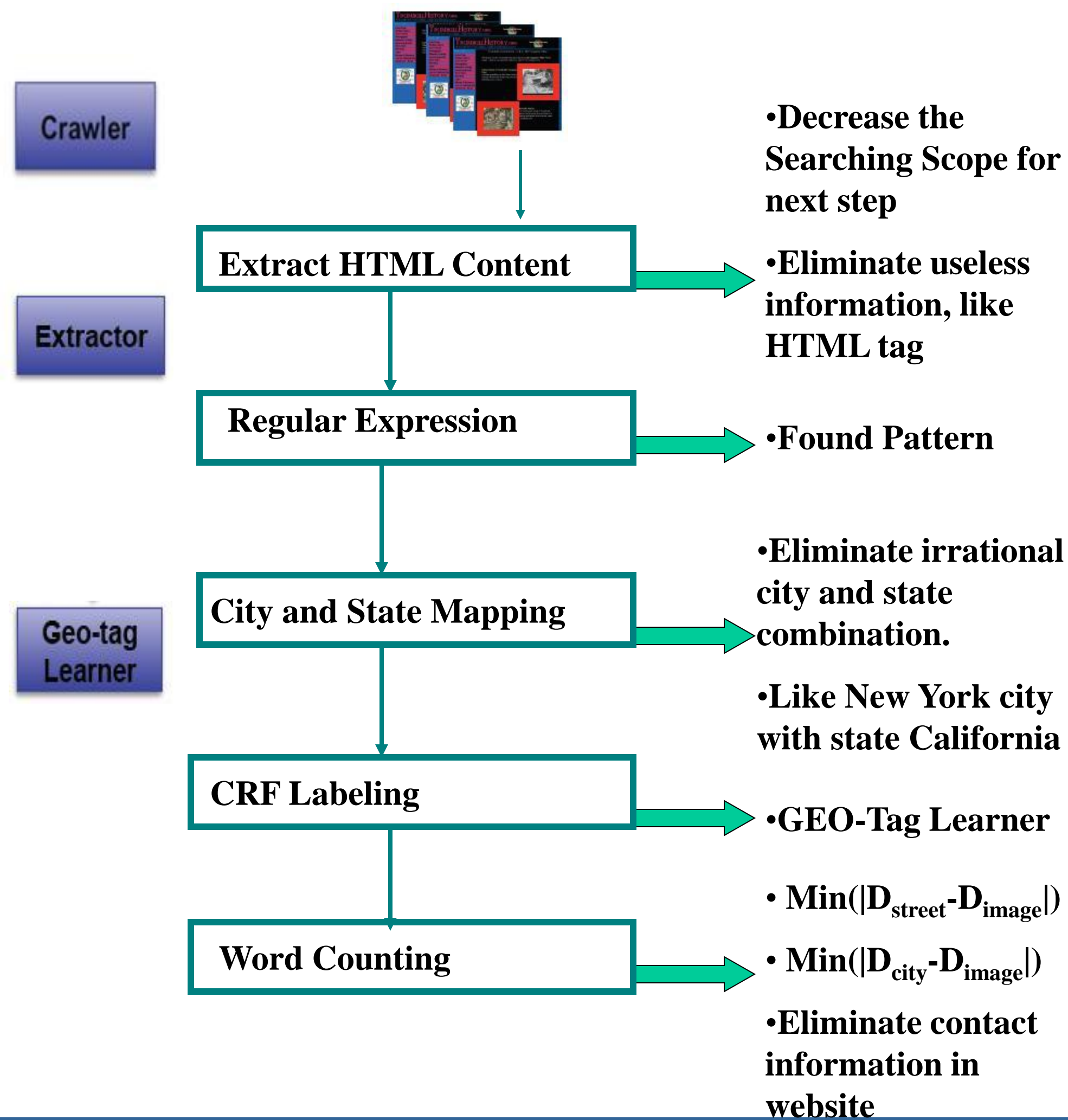
Abstract

With the internet technology spreads to everywhere of the world, it is possible to find the photographs for the same location or find the current photograph based on an old photograph. It is not difficult for human beings to do it manually. Our goal is how to do it automatically. The whole automatic process includes extracting address from the website, and using CRF, Lesk Algorithm gets the most possible one. In our proposed approach, we divide all websites into 4 cases firstly as the chart shown below.

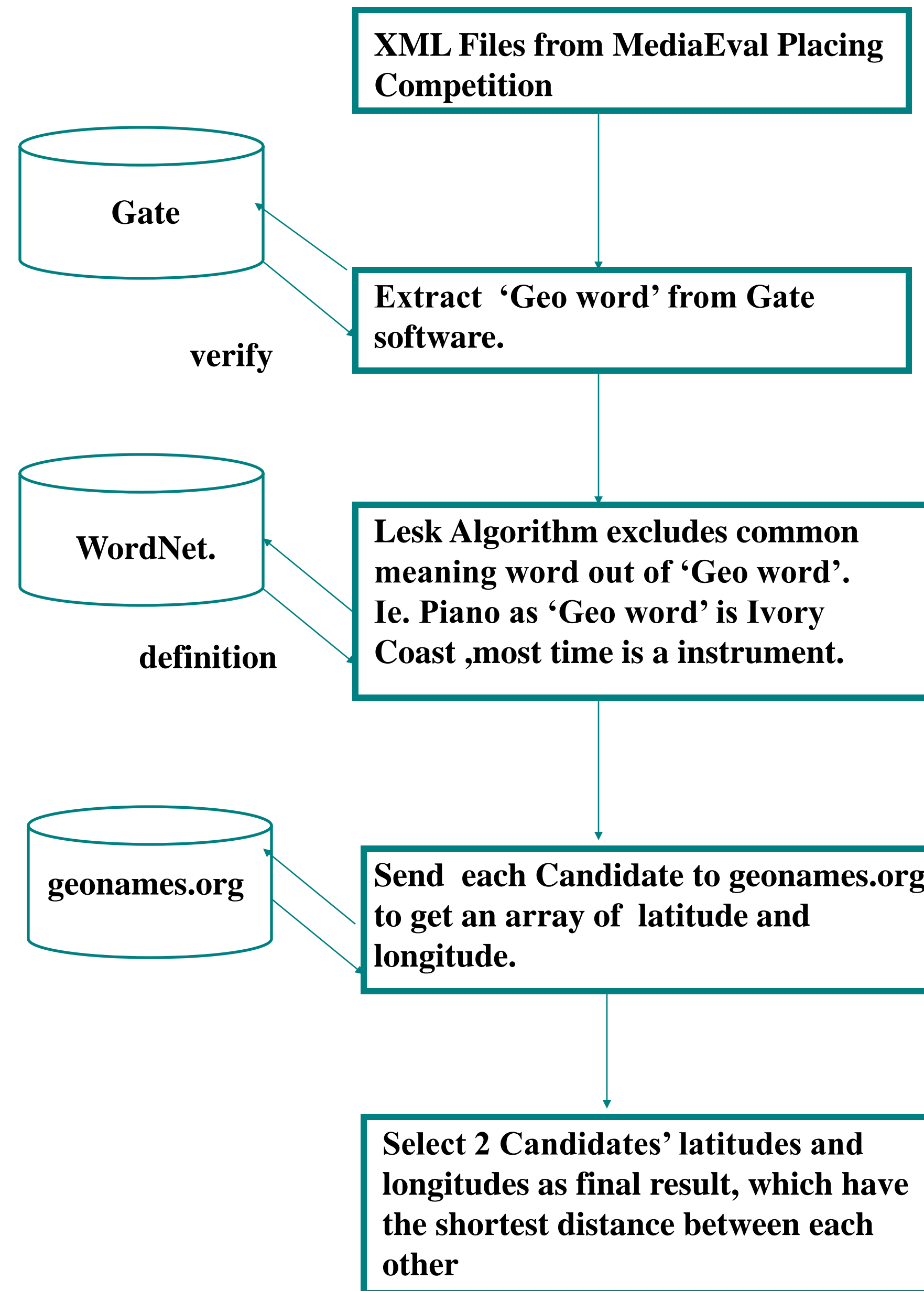
Case1	website that can provide exact address.	ie. 8 park ave, Bridgeport, ct
Case2	website that can strong clue about the current address	ie. Brooklyn, New York City
Case3	the website that can provide weak clue about the address	ie. New York City
Case4	the website that provides no clue about current address.	ie. Singing with piano

In this poster, we will solve Case1, Case2 and Case3.

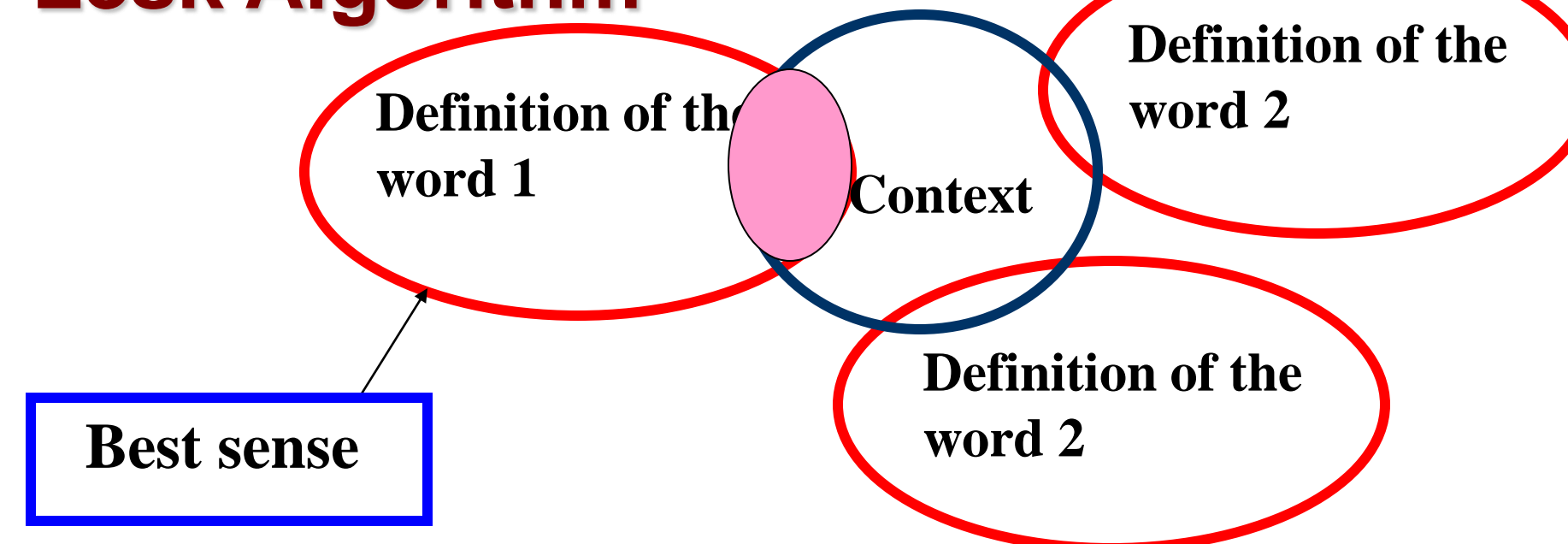
Case1 Approach



Case2 and Case3 Approach



Lesk Algorithm

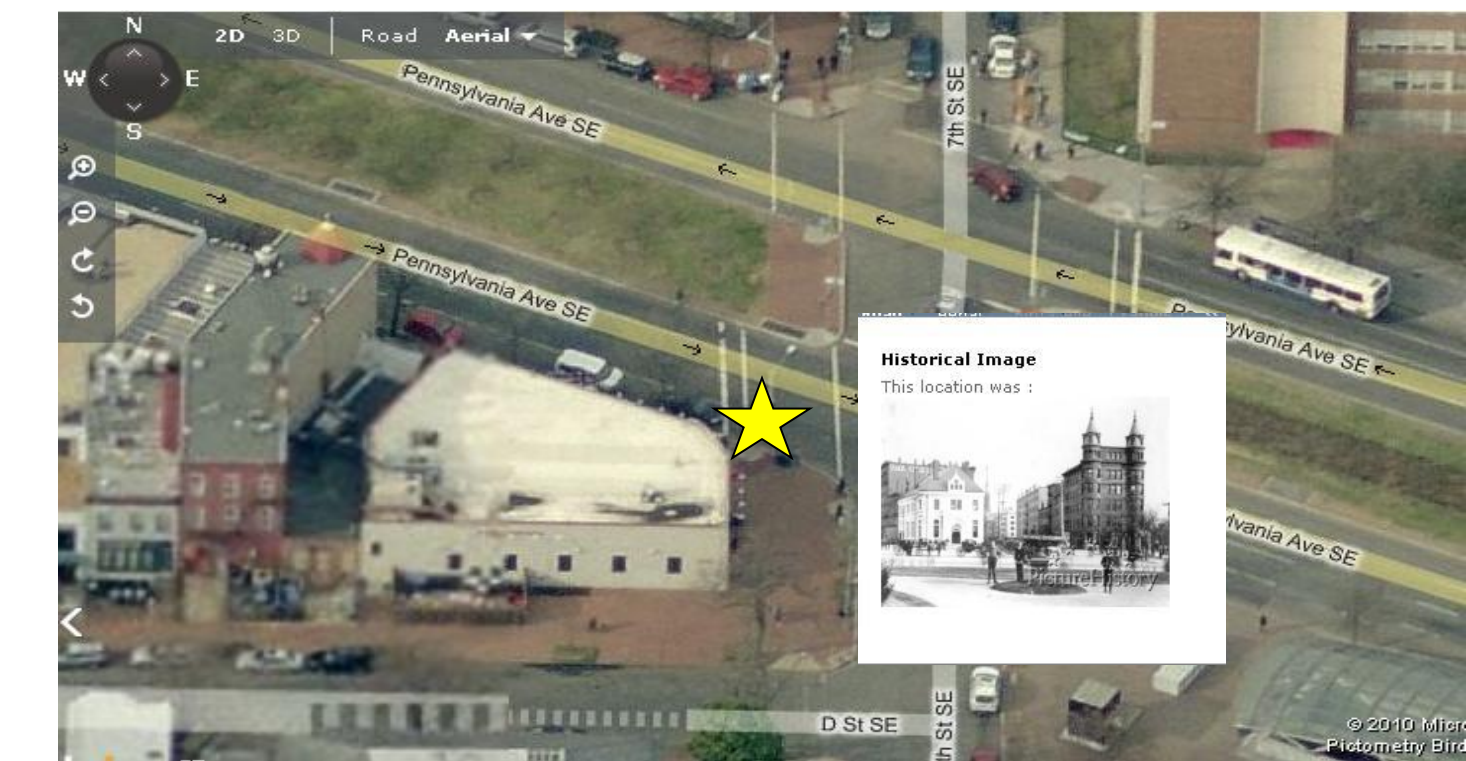


Results

Case 1 result

Total Links	Street Accuracy	City Accuracy	State Accuracy	Cardinality Type
100	89%	92%	94%	1:1

Case 1 sample



Case 2 and Case3 result

Total XML files	Total files with geo word	Average Distance	Total files accuracy<=10km	Total files accuracy<=100km
1856	1491	1720.774KM	603	864

Note: In case2, MediaEval placing competition provides the ground truth to compare our results. Here all distance and accuracy refer to the distance between the point we found and the corresponding ground truth.

CONCLUSION

Our goal is to find the geographical information from a website automatically. We analyze the common features of the websites and divide them into different cases. Outside geographical database are used for helping geographical information extraction. Then, we used CRF and Lesk algorithm separately for Case1, Case2 and Case3 to do word sense disambiguation. Our future plan is to establish pattern database and exploit web crawler ability to improve accuracy.