Automatic Extraction and Geo-spatial Visualization of FEMA National Situation Updates

Chi-Chun Pan and Prasenjit Mitra

College of Information Sciences and Technology, The Pennsylvania State University, University Park, PA, 16802.

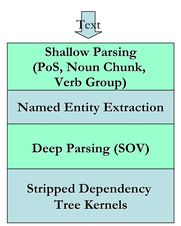
Introduction

- Rapid responsiveness is the key of successful crisis management. However, emergency situations may Come with an overload of information which is difficult to be processed in manual faction. Hence, automatically extracting useful pieces of information from vast sources of textual data is vital for such scenarios.
- In this project, we fuse information extraction and text mining techniques for **Visual Analytics**. We design a hybrid information extraction system to extract spatial-temporal information from text documents using both rule-based and machine-learning methods.
- A prototype system has been deployed to visualize FEMA National Situation Updates. Daily reports are fetched from the FEMA website and split into several incidents. Each incident then will be classified into topics based on word frequency and tagged with location names. The extracted information is stored in a repository and can be visualized with GIS applications such as Google Map. The system provides an intuitive way to browse and visualize FEMA situation updates.

Research Problems

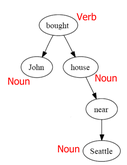
- How to extract and manage information based on interestingness?
  - Named Entities represent key concepts in a text. Extracting Named Entities could provide an abstract view of the whole document.
- How to find meaningful connection between pieces of information?
  - Identifying relationships between Named Entities with both syntactic and semantic meanings.
- How to ground textual data to a map?
  - Finding location entities within a text documents and resolving toponym ambiguity.
- How to analyze a vast amount of information for time critical decisions?
  - **Visual Analytics** provides an intuitive approach to improve understanding of the data being processed.
- Those research problems lead to a spatial-temporal information system for supporting analytical reasoning.

Methods



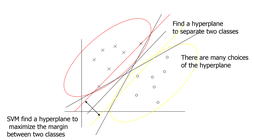
Text Processing Flow

Relation Extraction
- We define two level of relations
- Syntactic Relations
  - Syntactical parsing and rule-based system to identify SVO structure.
  - Find a subtree rooted in a verb which connects two entities
  - Between Named Entities
    - (John, bought, Seattle)
  - If extracting common nouns
    - (John, bought, house)

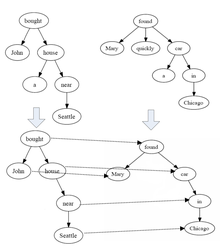


Syntactic Relations

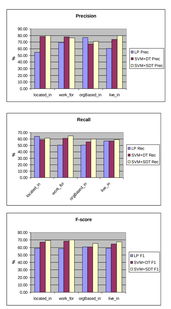
- Semantic Relations
  - Machine-Learning based system using Support Vector Machine (SVM)
  - Model trained for
    - Located_at (Person, Location)
    - Located_in (Location, Location)
    - OrgBased_on (Organization, Location)
    - Work_for (Person, Location)
  - Could be extended by providing richer training sets



Support Vector Machine

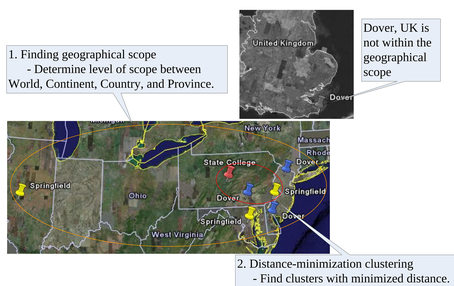


Semantic Relations

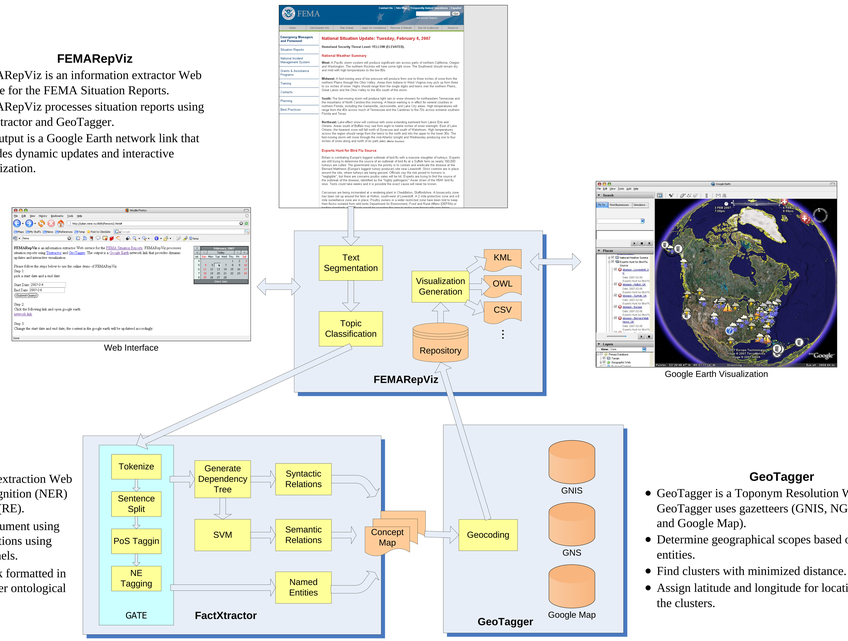


Performance for Relation Extracion

Geocoding



Architecture



Web Interface

Google Earth Visualization

FEMARepViz

- FEMARepViz is an information extractor Web service for the FEMA Situation Reports.
- FEMARepViz processes situation reports using FactXtractor and GeoTagger.
- The output is a Google Earth network link that provides dynamic updates and interactive visualization.

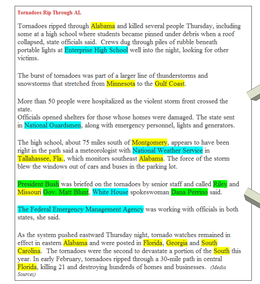
FactXtractor

- FactXtractor is an information extraction Web service for Named Entity Recognition (NER) and Entity Relation Extraction (RE).
- FactXtractor processes text document using GATE and indentify entity relations using Stripped Dependency Tree kernels.
- The output is a concept network formatted in OWL and can be used for further ontological reasoning.

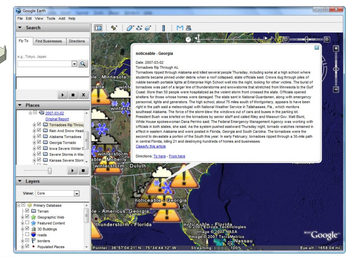
GeoTagger

- GeoTagger is a Toponym Resolution Web service. GeoTagger uses gazetteers (GNIS, NGA GEOnet, and Google Map).
- Determine geographical scopes based on location entities.
- Find clusters with minimized distance.
- Assign latitude and longitude for location entities in the clusters.

Results



- Each situation extracted from Web pages will be tagged.
- Concept maps can be created with Named Entities and Entity Relations.
- Situations can be geotagged and pintpointed onto Google Earth.



- Each situation is tagged with a timestamp.
- Use Google Earth to create geo-temporal visualizations to supporting analytical reasoning.

Conclusion

- We design an information extraction system for Visual Analytics.
- We propose the Stripped Dependency Tree Kernels and a heuristic-based geocoding algorithm.
- Applications include emergency situation patterns analysis and real-time emergency updates monitoring.

Future Work

- Design better algorithms to improve the performance for relation extraction and geocoding.
- Aggregate disparate data sources to provide a world-wide real-time emergency monitoring.
- Provide statistical analysis module for geo-temporal data.

References

[1] M. Collins and N. Duffy. Convolution kernels for natural language. In NIPS 2001, 2001.
[2] C. Cortes and V. Vapnik. Support-vector networks. Machine Learning, 20(3):273–297, 1995.
[3] A. Culotta and J. Sorensen. Dependency tree kernels for relation extraction. In ACL, 2004.
[4] S. M. Harabagiu, C. A. Bejan, and P. Morarescu. Shallow semantics for relation extraction. In Proceedings of the Nineteenth International Joint Conference on Artificial Intelligence (IJCAI '05), pages 1061–1066, Edinburgh, Scotland, UK, 2005.
[5] F. Peng, D. Schuurmans, and S. Wang. Language and task independent text categorization with simple language models. In NAACL '03: Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology, pages 110–117, Morristown, NJ, USA, 2003. Association for Computational Linguistics.
[6] D. Roth and W.-T. Wih. A linear programming formulation for global inference in natural language tasks. In Proceedings of CoNLL-2004, 2004.
[7] D. Zelenko, C. Aone, and A. Richardella. Kernel methods for relation extraction. Journal of Machine Learning Research, 3:1083–1106, 2003.