

FactXtractor

Automated named entities and entity relationship extraction from text documents

Benefit: FactXtractor is an information extraction Web service that provides core technologies to support textual data analysis. It extracts named entities and the relations between named entities from text. When combined with visualization tools and visual interfaces, FactXtractor can help users understand the relationships among entities of large document collections efficiently. FactXtractor forms the basis for several search and knowledge management tools that can be built on top of it.

Given a set of documents, FactXtractor extracts relationships among named entities within the text. FactXtractor performs relationship extraction on two levels. It extracts terms that describe relationships among named entities directly from the sentences in which the named entities appear. Secondly, FactXtractor uses machine learning to learn to extract positive examples of relationships of interest from text. For example, if we want to extract the relationship “located_in”, FactXtractor would identify that “situated at”, “headquartered at”, etc., are synonyms of that relation and extract all entities for which the relationship holds.

Extraction of named entities and relationships is a useful step toward understanding the subjects of a document and their actions. Consequently, an event-detection tool can be built that uses FactXtractor to identify the actors in an event and their relationships. Named entity extraction has also been shown to improve the results of search engines. Our FEMA situation update report-parsing and visualization tool, FemaRepViz, utilizes FactXtractor as a base tool to extract the locations named in the documents, which are then plotted using icons on a map. FactXtractor utilizes the open-source named entity extraction tool, GATE, and improves the precision of the named entity extraction using additional rules. In empirical evaluations, FactXtractor has been shown to have acceptable accuracy. A web-service interface allows any external tool to interact with FactXtractor to submit a set of documents containing named entities (and relationships among them) that need to be extracted.

Tornadoes Rip Through AL

Tornadoes ripped through Alabama and killed several people Thursday, including some at a high school where students became pinned under debris when a roof collapsed, state officials said. Crews dug through piles of rubble beneath portable lights at Enterprise High School well into the night, looking for other victims.

The burst of tornadoes was part of a larger line of thunderstorms and snowstorms that stretched from Minnesota to the Gulf Coast.

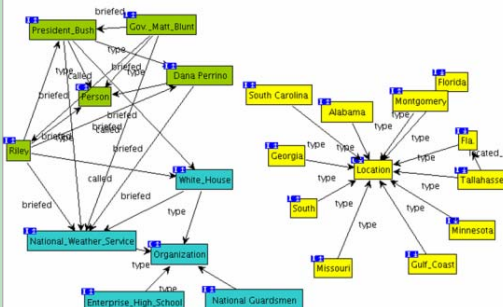
More than 50 people were hospitalized as the violent storm front crossed the state. Officials opened shelters for those whose homes were damaged. The state sent in National Guardsmen, along with emergency personnel, lights and generators.

The high school, about 75 miles south of Montgomery, appears to have been right in the path said a meteorologist with National Weather Service in Tallahassee, Fla., which monitors southeast Alabama. The force of the storm blew the windows out of cars and buses in the parking lot.

President Bush was briefed on the tornadoes by senior staff and called Riley and Missouri Gov. Matt Blunt, White House spokeswoman Jana Perrino said.

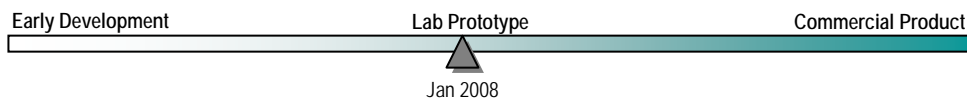
The Federal Emergency Management Agency was working with officials in both states, she said.

As the system pushed eastward Thursday night, tornado watches remained in effect in eastern Alabama and were posted in Florida, Georgia and South Carolina. The tornadoes were the second to devastate a portion of the South this year. In early February, tornadoes ripped through a 30-mile path in central Florida, killing 21 and destroying hundreds of homes and businesses. (Media Sources)



Results of a processed text document and the concept map populated by FactXtractor:
The concept map is visualized in ConceptVISTA. Each named entity is colored according to its entity type (person, place, or organization). The links between entities represent the relationship identified by FactXtractor.

Funded by: Department of Homeland Security.



For more information, contact:
Alan MacEachren, (814) 865-7491, maceachren@psu.edu
www.geovista.psu.edu/NEVAC