

Extracting and Combining Information from Text Sources

Eduard Hovy
USC Information Sciences Institute

The availability of massive amounts of information on the web means that a lot of the information is repeated, allowing automated systems to learn how to identify and extract information of interest to the analyst. By combining information in various ways, interesting complex concepts and patterns can be identified. I will outline projects underway at our USC center that build biographies, identify psychological states, extract biomedical information, and determine the social relationship between people, and I will discuss both the actions required of the system user (as teacher) and the shortcomings of modern information extraction techniques.