

Building a Dynamic News Category Network for News Sources Recommendations

Authors : Swati Gupta, Shagun Sodhani, Dhaval Patel, Biplab Banerjee

Abstract : It is generic that news sources publish news in different broad categories. These categories can either be generic such as Business, Sports, etc. or time-specific such as World Cup 2015 and Nepal Earthquake or both. It is up to the news agencies to build the categories. Extracting news categories automatically from numerous online news sources is expected to be helpful in many applications including news source recommendations and time specific news category extraction. To address this issue, existing systems like DMOZ directory and Yahoo directory are mostly considered though they are mostly human annotated and do not consider the time dynamism of categories of news websites. As a remedy, we propose an approach to automatically extract news category URLs from news websites in this paper. News category URL is a link which points to a category in news websites. We use the news category URL as a prior knowledge to develop a news source recommendation system which contains news sources listed in various categories in order of ranking. In addition, we also propose an approach to rank numerous news sources in different categories using various parameters like Traffic Based Website Importance, Social media Analysis and Category Wise Article Freshness. Experimental results on category URLs captured from GDELT project during April 2016 to December 2016 show the adequacy of the proposed method.

Keywords : news category, category network, news sources, ranking

Conference Title : ICSIR 2017 : International Conference on Semantic Information Retrieval

Conference Location : Paris, France

Conference Dates : July 20-21, 2017