

Thesis Title *

Kno.e.sis Center, CSE Department
Wright State University, Dayton, OH - USA
{pramod }@knoesis.org

Abstract. In this era of big-data, collecting relevant data is the major first step towards analyzing, let alone get some information of it. There has been a lot of research on analysis part of it yet a very little is done in the data collection part. Traditional way of collecting/crawling data, from social networking sites say Twitter, include manually coming up with the keywords for an event and start crawling. This has its own disadvantages

1) It is very difficult to come up with all the relevant keywords for an event manually. Also the keywords for an event are platform dependent (For example the keywords used in twitter may be totally different from that of G+ or Facebook). So we need a more efficient way.

2) Even though if we can come up with keywords, data on web changes every minute, events evolve over time and people may not be using the same keywords as before in-fact they may be using new keywords. Manually keeping up with this is very difficult rather say not possible.

So there is a necessity of an efficient and automatic way of doing this. In this thesis, we develop a mechanism by which the system, given an event, will automatically get the keywords for crawling and also updates the keywords as the event evolves. Our algorithm uses both Semantic Web and Information Retrieval techniques to come up with the keywords and to dynamically/automatically update the keywords as the event evolves over time. Due to the informal nature of the content in social media, using IR techniques alone won't give us correct results, so we combine them with Semantic Web Techniques (using Wikipedia and DBPedia to get semantically related entities) to extract the correct keywords. And to keep up with the evolution of the event we track revisions of Wikipedia page, updating the keywords when there is a change in the Wiki page. Finally we evaluate our system with

Keywords: Semantic Web, Social Web, Dynamic Events, Twitter

1 Introduction

* .