**LITERATURE SURVEY**

Literature [survey](http://www.blurtit.com/q876299.html) is the most important step in software development process. Before developing the tool it is necessary to determine the time factor, economy n company strength. Once these things r satisfied, ten next steps are to determine which operating system and language can be used for developing the tool. Once the [programmers](http://www.blurtit.com/q876299.html) start building the tool the programmers need lot of external support. This support can be obtained from senior programmers, from [book](http://www.blurtit.com/q876299.html) or from websites. Before building the system the above consideration are taken into account for developing the proposed system.

**G. Jain, M. Sharma, and B. Agarwal, ‘‘Spam detection in social media using convolutional and long short term memory neural network,**

As the use of the Internet is increasing, people are connected virtually using social media platforms such as text messages, Facebook, Twitter, etc. This has led to increase in the spread of unsolicited messages known as spam which is used for marketing, collecting personal information, or just to offend the people. Therefore, it is crucial to have a strong spam detection architecture that could prevent these types of messages. Spam detection in noisy platform such as Twitter is still a problem due to short text and high variability in the language used in social media. In this paper, we propose a novel deep learning architecture based on Convolutional Neural Network (CNN) and Long Short Term Neural Network (LSTM). The model is supported by introducing the semantic information in representation of the words with the help of knowledge-bases such as WordNet and ConceptNet. Use of these knowledge-bases improves the performance by providing better semantic vector representation of testing words which earlier were having random value due to not seen in the training. Proposed Experimental results on two benchmark datasets show the effectiveness of the proposed approach with respect to the accuracy and F1-score.

**F. Pierri and S. Ceri, ‘‘False news on social media: A data-driven survey,’’**

In the past few years, the research community has dedicated growing interest to the issue of false news circulating on social networks. The widespread attention on detecting and characterizing false news has been motivated by considerable backlashes of this threat against the real world. As a matter of fact, social media platforms exhibit peculiar characteristics, with respect to traditional news outlets, which have been particularly favorable to the proliferation of deceptive information. They also present unique challenges for all kind of potential interventions on the subject. As this issue becomes of global concern, it is also gaining more attention in academia. The aim of this survey is to offer a comprehensive study on the recent advances in terms of detection, characterization and mitigation of false news that propagate on social media, as well as the challenges and the open questions that await future research on the field. We use a data-driven approach, focusing on a classification of the features that are used in each study to characterize false information and on the datasets used for instructing classification methods. At the end of the survey, we highlight emerging approaches that look most promising for addressing false news.

**T. Wu, S. Wen, Y. Xiang, and W. Zhou, ‘‘Twitter spam detection: Survey of new approaches and comparative study,’’**

Twitter spam has long been a critical but difficult problem to be addressed. So far, researchers have proposed many detection and defence methods in order to protect Twitter users from spamming activities. Particularly in the last three years, many innovative methods have been developed, which have greatly improved the detection accuracy and efficiency compared to those which were proposed three years ago. Therefore, we are motivated to work out a new survey about Twitter spam detection techniques. This survey includes three parts: 1) A literature review on the state-of-art: this part provides detailed analysis (e.g. taxonomies and biases on feature selection) and discussion (e.g. pros and cons on each typical method); 2) Comparative studies: we will compare the performance of various typical methods on a universal testbed (i.e. same datasets and ground truths) to provide a quantitative understanding of current methods; 3) Open issues: the final part is to summarise the unsolved challenges in current Twitter spam detection techniques. Solutions to these open issues are of great significance to both academia and industries. Readers of this survey may include those who do or do not have expertise in this area and those who are looking for deep understanding of this field in order to develop new methods.

**M. Hussain, M. Ahmed, H. A. Khattak, M. Imran, A. Khan, S. Din, A. Ahmad, G. Jeon, and A. G. Reddy, ‘‘Towards ontology-based multilingual URL filtering: A big data problem,’**

Web content filtering is one among many techniques to limit the exposure of selective content on the Internet. It has gotten trivial with time, yet filtering of multilingual web content is still a difficult task, especially while considering big data landscape. The enormity of data increases the challenge of developing an effective content filtering system that can work in real time. There are several systems which can filter the URLs based on artificial intelligence techniques to identify the site with objectionable content. Most of these systems classify the URLs only in the English language. These systems either fail to respond when multilingual URLs are processed, or over-blocking is experienced. This paper introduces a filtering system that can classify multilingual URLs based on predefined criteria for URL, title, and metadata of a web page. Ontological approaches along with local multilingual dictionaries are used as the knowledge base to facilitate the challenging task of blocking URLs not meeting the filtering criteria. The proposed work shows high accuracy in classifying multilingual URLs into two categories, white and black. Evaluation results conducted on a large dataset show that the proposed system achieves promising accuracy, which is on a par with those achieved in state-of-the-art literature on semantic-based URL filtering.

**M. U. S. Khan, M. Ali, A. Abbas, S. U. Khan, and A. Y. Zomaya, ‘‘Segregating spammers and unsolicited bloggers from genuine experts on Twitter,’**

Online Social Networks (OSNs) have not only significantly reformed the social interaction pattern but have also emerged as an effective platform for recommendation of services and products. The upswing in use of the OSNs has also witnessed growth in unwanted activities on social media. On the one hand, the spammers on social media can be a high risk towards the security of legitimate users and on the other hand some of the legitimate users, such as bloggers can pollute the results of recommendation systems that work alongside the OSNs. The polluted results of recommendation systems can be precarious to the masses that track recommendations. Therefore, it is necessary to segregate such type of users from the genuine experts. We propose a framework that separates the spammers and unsolicited bloggers from the genuine experts of a specific domain. The proposed approach employs modified Hyperlink Induced Topic Search (HITS) to separate the unsolicited bloggers from the experts on Twitter on the basis of tweets. The approach considers domain specific keywords in the tweets and several tweet characteristics to identify the unsolicited bloggers. Experimental results demonstrate the effectiveness of the proposed methodology as compared to several state-of-the-art approaches and classifiers.