Annotated Bibliography

* Can social media be reliably analysed to provide useful statistical data?

Nadkarni, Prakash M; Ohno-Machado, Lucila; Chapman, Wendy W. (2011) Natural language processing: an introduction. Journal of the American Medical Informatics Association. [Online] 18(5), pp.544 - 551 [Accessed 19 October 2014] Available at:

<http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3168328/?tool=pmcentrez&report=abstract>

The paper gives a history of natural language processing(NLP), summarises NLP sub-issues, highlights a selection of NLP being used in the medical field, briefly describes some machine-learning techniques used for NLP, describes how newer NLP systems are designed with examples and concludes on whether NLP applications will become a commodity and how NLP would benefit from this. These claims are precise because of the wealth of references that are used to back themselves up, these references are mainly from around the same time period as the papers publication and come from reputable institutions and journal, therefore they act as their evidence and ensures the papers trustworthiness. The paper only gives a brief description with figures of the techniques/algorithms involved but does serve to introduce and give a basic understanding of what it covers.

Kumar, S., Hu, X., & Liu, H. (2014). A behavior analytics approach to identifying tweets from crisis regions. Cover Art: In Proceedings of the 25th ACM conference on Hypertext and social media - HT ’14 [Online]. Santiago, Chile 1-4 September. New York: ACM. [Accessed 6 November 2014]. Available at: <http://dl.acm.org/citation.cfm?id=2631775.2631814>

The paper discusses the issue of gathering tweets from crisis regions only so their analysis can be increasingly efficient as well as to tell whether a tweet is from a crisis region from the tweets information and its owner. It concludes that their method is successful in this task and details the planned future work. The paper includes a wealth of reputable sources on the topics covered that are appropriately referenced as well as their own data from the testing of their method such as in graph form which backs up its conclusion.

Wang, A., Hoang, C., & Kan, M.-Y. (2013). Perspectives on crowdsourcing annotations for natural language processing. Language Resources & Evaluation, 47(1), pp.9–31 [Accessed 7 November 2014]. Available at: <http://wk6kg9sd8m.search.serialssolutions.com/?ctx\_ver=Z39.882004&ctx\_enc=info%3Aofi%2Fenc%3AUTF8&rfr\_id=info:sid/summon.serialssolutions.com&rft\_val\_fmt=info:ofi/fmt:kev:mtx:journal&rft.genre=article&rft.atitle=Perspectives+on+crowdsourcing+annotations+for+natural+language+processing&rft.jtitle=Language+Resources+and+Evaluation&rft.au=Wang%2C+Aobo&rft.au=Hoang%2C+Cong+Duy+Vu&rft.au=Kan%2C+MinYen&rft.date=20130301&rft.pub=Springer+Netherlands&rft.issn=1574020X&rft.eissn=15740218&rft.volume=47&rft.issue=1&rft.spage=9&rft.epage=31&rft\_id=info:doi/10.1007%2Fs10579-012-9176-1&rft.externalDBID=n%2Fa&rft.externalDocID=2013\_10579\_47\_1

\_9176&paramdict=en-US>

The source describes annotation programs which are used as a way for NLP to help machine learning algorithms by providing it with better or in some cases just larger amounts of data to learn from, crowdsourcing is shown as a new way to gain the data needed as input for machine learning as well as how it affects future annotation programs, finally it concludes that humans and computers can work together to solve problems where computers cannot do the task themselves and that while research into this area is relatively new it shows promise. It is backed up in its bibliography by reputable sources of information on the topics covered which are at the forefront of current research as well as quantitative evidence of their testing of several applications designed for the task at hand.