|  |  |  |  |
| --- | --- | --- | --- |
| **Method** | **Description** | **Example** | **Limitations** |
| Sentiment Analysis | Predicting the mood of the sentence like whether it is positive or negative [1]. | *“Product XYZ is good but expensive”*  The above statement states two aspects of the product XYZ where “product is good” shows the positive or favorable statement and “product XYZ is expensive” shows the negative or unfavorable statement [3]. | 1. Context – A decision cannot be made based on the words that are used in the context, as there will be two different meaning for the same word [5]. 2. Regional Variations – Language used in the context is a major limitation as different language words have different meaning [5]. |
| Optical Character Recognition in Natural Language Processing | Extracting useful information from the given image [1]. | Optical Character Recognition is used in Banking sectors to process the checks, just a scan of the check will process the transaction without any human involvement and it is also used in other industries like finance, education and legal industries to digitize the records or documents [4].  OCR is especially used to extract information from the unstructured data as the data can be in any format like image, text, graphics etc.  In OCR the text has been scanned, preprocessed, segmented and data has been extracted. | 1. The main limitation in extracting information is because the data is often mixed with text and graphics [2]. 2. Variations in style and shape of the data [2]. 3. Variations because of subscripts and superscripts in the data [2]. |
| Data Mining | Extracting useful information from the huge database process is known as data mining [6]. | Data mining is applied in Healthcare Industry for the evaluation of treatment effectiveness, management of healthcare, customer relationship management and detection of fraud and abuse [7].  For instance, the outcomes of patients treated with different medicines for the same disease to determine which drug is best and cost effective [7]. | 1. Issues may arise with the missing, corrupted, inconsistent data as the information recorded will be in different format from different sources [7]. |
| Web Crawling in Information Retrieval | Automatic script or program that can download the contents from the web pages [6]. | Web crawlers are used as a service to collect the web pages periodically and archive it for future purposes [8]. | 1. Content selection is very poor sometimes it bypasses the low-quality content, irrelevant and redundant content as all the content from the web should be selected invariantly either the good or poor [8]. 2. Some content suppliers inject unwanted content to the site which results in misdirection [8]. |
| Text Analytics | Text analytics helps in retrieving the valuable text information from the unstructured and semi-structured data. Text analytics is also referred as text mining [9]. | Text analytics or text mining is used in many fields like Publishing and media, Banks, Telecommunications etc. [11].  In the banking sectors CRM applications are more common and aimed at improving the management of customer communication, by automatic systems of message re-routing and with applications supporting the search engines asking questions in natural language [11]. | 1. Lot of software programming is needed to extract textual information from different sources [9]. 2. Managing the unstructured data from various sources is complicated [9]. |
| Big Data Analytics | Big data refers to large or complex data that are difficult to manage with the traditional software, hardware and data management tools [10]. | Big data analytics is used in many industries. For instance, in the health care industry it is used to detect the disease at the earlier stage where they can be treated more easily and effectively [10]. | 1. Data extraction and cleaning is one of the limitation in big data analytics. 2. Data Integration is another limitation as the data will be coming from different databases or web portals. |
| Auto completion in NLP | Predicting the word which follow the sequence of words [1].  For Example, the word “Blue” is more likely to follow the sequence “The sky is”. |  |  |
| Part-of-speech tagging | Tagging the part of speech for each word such as noun, verb, adverb, etc. [1]. |  |  |

**REFERENCES**

1. [Gabriele Bavota](http://ieeexplore.ieee.org.ezproxy.gannon.edu/search/searchresult.jsp?searchWithin=%22Authors%22:.QT.Gabriele%20Bavota.QT.&newsearch=true), Mining Unstructured Data in Software Repositories: Current and Future Trends, [Software Analysis, Evolution, and Reengineering (SANER), 2016 IEEE 23rd International Conference on](http://ieeexplore.ieee.org.ezproxy.gannon.edu/xpl/mostRecentIssue.jsp?punumber=7476313) 14-18 March 2016

# Line Eikvil, Optical Character recognition, December 1993

1. Tetsuya Nasukawa, Jeonghee Yi, Sentiment Analysis: Capturing Favorability using Natural Language Processing
2. Ravina Mithe, Supriya Indalkar, Nilam Divekar, Optical Character Recognition, International Journal of Recent technology and Engineering, March 2013

# [V Nagarjana Devi Duvvuri](http://ieeexplore.ieee.org.ezproxy.gannon.edu/search/searchresult.jsp?searchWithin=%22Authors%22:.QT.V%20Nagarjana%20Devi%20Duvvuri.QT.&newsearch=true), [V.Rajanikanth Thatiparthi](http://ieeexplore.ieee.org.ezproxy.gannon.edu/search/searchresult.jsp?searchWithin=%22Authors%22:.QT.V.Rajanikanth%20Thatiparthi.QT.&newsearch=true), [Rajashekar Pantangi](http://ieeexplore.ieee.org.ezproxy.gannon.edu/search/searchresult.jsp?searchWithin=%22Authors%22:.QT.Rajashekar%20Pantangi.QT.&newsearch=true), [Akhil Gangavarapu](http://ieeexplore.ieee.org.ezproxy.gannon.edu/search/searchresult.jsp?searchWithin=%22Authors%22:.QT.Akhil%20Gangavarapu.QT.&newsearch=true), Sentiment Analysis Using Harn Algorithm, [IT Convergence and Security (ICITCS), 2016 6th International Conference on](http://ieeexplore.ieee.org.ezproxy.gannon.edu/xpl/mostRecentIssue.jsp?punumber=7740253) 26-26 Sept. 2016

# [Chandni Saini](http://ieeexplore.ieee.org.ezproxy.gannon.edu/search/searchresult.jsp?searchWithin=%22Authors%22:.QT.Chandni%20Saini.QT.&newsearch=true), [Vinay Arora](http://ieeexplore.ieee.org.ezproxy.gannon.edu/search/searchresult.jsp?searchWithin=%22Authors%22:.QT.Vinay%20Arora.QT.&newsearch=true), Information retrieval in web crawling: A survey,  [Advances in Computing, Communications and Informatics (ICACCI), 2016 International Conference on](http://ieeexplore.ieee.org.ezproxy.gannon.edu/xpl/mostRecentIssue.jsp?punumber=7592392) 21-24 Sept 2016

# Hian Chye Koh, Gerald Tan, Data Mining Applications in Healthcare

# Namrata H.S, B.S. Satpute, Pramod Patil, Web Forum Crawling Techniques, International Journal of Computer Applications (0975 – 8887) Volume 85 – No 17, January 2014

# Kalli Srinivasa Nageswara Prasad, Prof S. Ramakrishna, Text Analytics to Data Warehousing.

# Wullianallur Raghupathi and Viju Raghupathi, Big data analytics in healthcare: promise and potential, 07 Feb 2014

# Vishal Gupta, Gurpreet S.Lehal, A survey of Text Mining Techniques and Applications, Jornal of Emerging Technologies in Web Intelligence, August 2009

# Saumya Salian, Challenges with Big Data Analytics, International Journal of Science and Research