Abstract:

Ontology is an effective and efficient way to understand a concept and any and everything related to it. It is a great tool in the world of deep learning and machine learning, where a system is trained to understand, interpret, analyze and derive results. The machine learning models built using Natural Language Processing (NLP) techniques resulting in knowledge graphs have evolved over the years making it applicable on text data pertaining to any field possible, containing information in various text formats. The world of medicine is no different and the NLP techniques have facilitated the understanding of the existing research and its evolution over the years. With focus on breast cancer, we shall develop a machine-learning model using NLP techniques that will identify the various aspects of breast cancer such as causes, symptoms, diagnosis, treatment and/or its side effects.

Keywords:

ontology, machine learning, deep learning, Natural Language Processing (NLP), breast cancer

Introduction:

With an increase in the number of diseases emerging and their impact on human well-being, there has been an increasing demand for faster diagnosis to ensure timely treatments. There is an enormous amount of bio-medical data available today. The world of big data, machine learning and deep learning is a great source to analyze this data and obtain results.

Availability of research work in the field of bio-medical analyzing and understanding the causes, symptoms, diagnosis, cure and effects of various diseases has eased the process of finding cures for newly evolving diseases in a better, faster way. All these information present in various forms of text data may be analyzed to identify various ontologies helping one understand, analyze and fight these diseases. The natural language processing (NLP) techniques available in the world of deep learning and machine learning process to be an efficient and effective means of analyzing the information (data) available to identify solutions.

Related Work:

There exists a tremendous amount of research in the area of ontology creation in bio-medical domain. A prototype system has been proposed using the approach of semantic technology leveraging the staging manual's automatic parsing of data. Further, additional biomarkers were included when developing the cancer staging manual. Development of the ontology involved the use of generic terms used within the community to map the terms associated with breast cancer.

Proposed Work:

We propose to build a machine leaning model using NLP techniques helping with the identification of causes, symptoms, diagnosis, treatment and/or side-effects related to breast cancer. The model obtained shall be used to build ontologies relating to various aspects of the disease.

The process involves the following steps.

1. Data Collection/Extraction
2. Data Pre-processing
3. Information Extraction
4. Information Retrieval

5. Ontology Generation

Data Set Used:

The data used to build are model has been obtained by collecting abstracts from several research papers available from the research done on breast cancer and its various aspects.

Collection of abstracts from several breast cancer related research papers

Implementation and Evaluation:

Will be updated as the work progresses.

Conclusion:

Will be updated as the work is completed.

Reference:

1. Links to various reference materials and research papers used as input data
2. https://arxiv.org/pdf/1807.07991.pdf
3. Complete references will be available upon completion of the project