International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878 (Online), Volume-8 Issue-6, March 2020

Prediction and Analysis of Extracting Relations using Spacy Model

G. Suganya, R. Porkodi

Abstract*: Text is an extremely rich resources of information. Each and every second, minutes, peoples are sending or receiving hundreds of millions of data. There are various tasks involved in NLP are machine learning, information extraction, information retrieval, automatic text summarization, question-answered system, parsing, sentiment analysis, natural language understanding and natural language generation. The information extraction is an important task which is used to find the structured information from unstructured or semi-structured text. The paper presents a methodology for extracting the relations of biomedical entities using spacy. The framework consists of following phases such as data creation, load and converting the data into spacy object, preprocessing, define the pattern and extract the relations. The dataset is downloaded from NCBI database which contains only the sentences. The created model evaluated with performance measures like precision, recall and f-measure. The model achieved 87% of accuracy in retrieving of entities relation.

***Keywords:* Text mining, Information Extraction, Natural Language Processing, Spacy**

I. INTRODUCTION

Information extraction is a challenging task in the field of Natural language processing. The system is spread in many areas such as machine translation, question answering systems, extracting the entities and events, relation extraction, etc. [1]

Natural language processing is a subfield of machine learning in which deals with processing and analyzing the data and also generating human speech. Nowadays, there are many machines are available for determining the meaning of a string of text. In the data science, large dataset to be analyzed by human experts but it takes time to do that. Since, the NLP techniques helps to get understanding of text meaning. Example: sentiment analysis, advertisement matching, chatbots, voice assistants and other applications where machines need to understand as well as respond faster [2].

Information extraction is used to extract salient features about the specified entities which used to populate the database to get more structured data. Identification of entities and its relation plays a major role in information extraction. The main tasks involved information extraction are named entity recognition and relation extraction [1].

The paper is organized as follows: Section 2 discuss about the literature review for the identification of relations; Section 3 describes the methodology framework; Results and discussion in the Section 4. Finally, the conclusion is drawn in Section 5.

II. LITERATURE REVIEW

(Bhasuran, Subramanian, & Natarajan, 2018) identified gene relationships related to high altitude disease by using principle of gene co-occurrence statistics from literature as well as network analysis [3]. (Cui et al., 2018) proposed an algorithm for predicting effects of drug combinations on various diseases by integrating the data of disease-related gene expression profiles with drug related gene expression profiles [4].

(Gupta et al., 2018) describes an automated tool called Disease- Expression Relation Extraction from Text (DEXTER) for extracting information from literature on gene and microRNA expression present in disease and also extracted rich expression information in different disease related situations [5].

(Y. Chang, Chu, Su, & Chen, 2016) had proposed IPT kernel approach (interaction Pattern Tree- PIPE)for identifying the interactions between proteins presented in biomedical literature [6].

(Bravo, Piñero, Queralt-rosinach, Rautschka, & Furlong, 2015)developed BeFree system which is used to identify relationships between heterogeneous biomedical entities with a special focus on genes and the diseases associated with them [7].

(Bchir, 2015) proposed an approach to extract disease-drug relations using machine learning algorithm [8].(Xu & Wang, 2013) developed a pattern learning relationships extraction algorithm used to extract drug-disease pairs from biomedical literature [9].

(Jeongkyun Kim et al., 2013)proposed a method called DigSee to identify disease related genes that are involved in development of disease from medline abstracts [10]. (Chun et al., 2006) proposed a system to extract the disease-gene relations from Medline abstracts [11].

III. METHODOLOGY FRAMEWORK

The objective of the research work is to identify the relations based on the defined pattern. The methodology framework is shown in figure 1.

Manuscript received on February 10, 2020.
Revised Manuscript received on February 20, 2020.
Manuscript published on March 30, 2020.
* Correspondence Author
Suganya. G*, Department of Computer Science, Bharathiar University, Coimbatore, India. Email: suganya.cs@buc.edu.in
Porkodi R Department of Computer Science, Bharathiar University, Coimbatore, India. Email: porkodi_r76@buc.edu.in

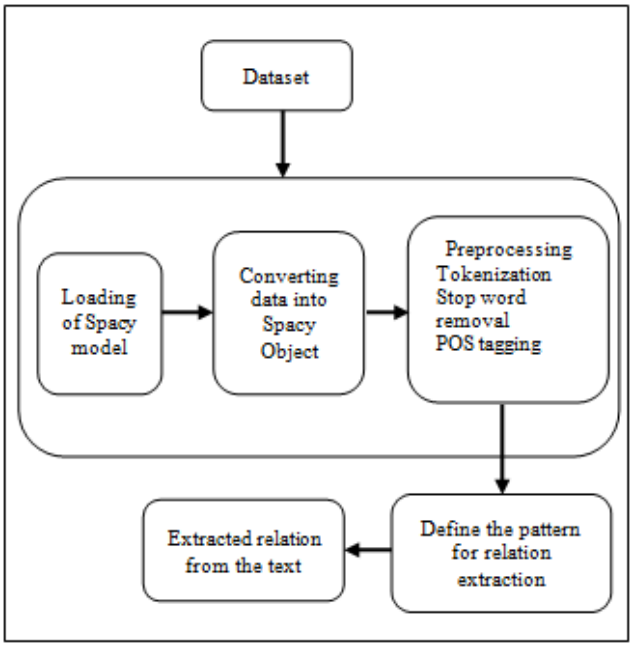

Figure 1. Methodology framework

Phase 1: Data modelling

The dataset is created by including of many sentences which contains only in the form of text.

Phase 2: Loading libraries

The needed library are pandas,math, nltk, numpy, re, spacy, matcher and displacy. The pandas package is used for processing of data, time series and statistics analysis and math is used for performing mathematical operations. The nltk is used for performing the preprocessing task and numpy is used for scientific operations. The re package is used for regular expression operations. The spacy is used for detects noun-phrases automatically and do the NLP tasks. The matcher is used for match sequences based on tokens and displacy package from spacy also imported which is used to view the named entity annotated sentence and visualization of object. The description about all the package is shown in table 1.

Phase 3: Preprocessing

Tokenizing the text: It is the process of splitting the text into pieces called tokens and ignore the special characters like punctuation marks, spaces and question marks, etc. The word tokenization is a critical step for many language processing of applications because often require input in the form of individual words rather than longer text. spacy tokenizer used which takes input in the form of Unicode text and outputs into sequence of token objects [12].

Cleaning text data: It helps us eliminate noise and distraction from text data. In spacy package, 312 stopwords are there. The stopwords are loaded default by includes Spacy [2].

Normalization / Lemmatization / Lemma: It is the one of cleaning process. It is a way of processing words that reduces them to their roots[2].

Phase 4: Entity detection

It is also called as entity recognition which is more advanced form of language processing that identifies important elements. Spacy package recognizes various types of named entities present in a document because models are statistical and depend on the training and testing data.

Part of Speech Tagging (POS): For performing POS tagging, first need to import en_core_web_sm model in spacy. Because the model contains two type of information such as dictionary and grammatical information required for analysis. Then load the model with load() function and identify the part of speech for all the word by using pos_[2].

Phase 5: Pattern definition and Extracting the relations

Define the pattern to extract relation from the input sentences. The pattern includes three components such as entity names, relation keyword and entity name which is default. The pattern may have more than three components.

Table 1. Package description

Library	Link	Description
pandas	https://pypi.org/project/pandas/	Data analysis, time series and statistics
math	https://www.tutorialsteacher.com/python/math-module	Performing mathematical operation
nltk	https://www.nltk.org/	text processing libraries for classification, tokenization, stemming, tagging, parsing, and semantic reasoning, wrappers for industrial-strength NLP libraries
numpy	https://numpy.org/	Scientific computing with python
re	https://docs.python.org/3/library/re.html	Regular expression matching operations
spacy	https://spacy.io/	Detect noun phrases
matcher	https://spacy.io/api/matcher	Match sequences of tokens based on pattern rules
displacy	https://spacy.io/usage/visualizers	Visualization dependencies and entities in a notebook.

IV. RESULT AND DISCUSSION

Dataset and Preprocessing Task

The dataset consists of totally 1000 sentences which is in the form of text and collected from PubMed Central Library. In that, 90:10 of the data as training and testing sentences. A snippet of the sample dataset is shown in below.

Example:

Sentence: CHD8 is activates the BRG1 which is associated with the SWI/SNF activate CHD7 data.

After loading the dataset, apply the pre- processing techniques such as tokenization, stopword removal and stemming. After done these tasks, the dataset will be in below form.

Example:

Sentence	After Pre-Processing
CHD8 is activates the BRG1 which is associated with the SWI/SNF activate CHD7 data.	CHD8 activates BRG1 associated SQI/SNP activate CHD7

After load the data into en_core_web_sm, convert or create the word vector by using #1. This shows the numeric representation of a word that communicates the relationship to other words.

Each and every word is interpreted as a unique and lengthy array of numbers. Its look like coordinates of two sets. The word vector is shown in figure 2.

#1

```
doc.vector
doc.vector.shape
```

There is no possible way to identify the meaning of the given meaning of the sentences, but the representation of word works well for all machines. Since the method allows to represent both meaning of word and other similar words.
Next, Apply the POS tag for the all the word by using pos_. A snippet of the POS tagging is shown in Example 1.

Example 1:

Sentence	POS tagging
CHD8 activates BRG1 associated SWI/SNF activate CHD7	CHD8 -->nsubj --> PROPN activates --> ROOT --> VERB BRG1 --> compound --> PROPN associated -->amod --> VERB SWI -->nmod --> PROPN / -->punct --> SYM SNF -->nsubj --> PROPN activate -->ccomp --> VERB CHD7 -->dobj --> PROPN

```
array(
[ 2.39400491e-01, -6.19609356e-01, -1.19236827e+00,  2.30014086e+00,
     -1.00512542e-01,  2.53967023e+00,  1.05286193e+00, -2.24507064e-01,
      1.13073695e+00,  2.53419638e+00,  1.35180342e+00,  5.68003833e-01,
      1.24180183e-01, -9.29246902e-01, -1.71437216e+00, -1.33210528e+00,
     -3.98938835e-01,  1.71908319e+00,  2.75532246e-01,  1.18906870e-01,
     -6.76979497e-02, -1.65830755e+00, -5.63504279e-01, -1.21358192e+00,
     -1.10092616e+00, -5.95269501e-01, -7.49114990e-01, -2.03692389e+00,
      3.38559210e-01, -1.77981186e+00, -2.21517742e-01, -1.11915207e+00,
      2.45217964e-01, -1.32969153e+00,  1.04475224e+00, -2.31239271e+00,
      1.61526775e+00, -4.73908275e-01, -3.18315059e-01, -8.78721178e-02,
      1.43729293e+00,  1.29211760e+00,  8.42701972e-01, -2.92787051e+00,
     -9.60983694e-01,  5.83246946e-01, -4.85554159e-01, -5.43210626e-01,
      5.29976487e-01,  1.91518641e+00, -5.32110810e-01, -1.70853758e+00,
      7.87986994e-01, -9.24992681e-01, -2.41718054e+00,  1.51324165e+00,
      4.06864375e-01,  2.34363317e-01,  3.14445823e-01,  2.90273637e-01,
      5.90195000e-01,  5.42960584e-01,  7.13709891e-01,  7.64942691e-02,
      2.40207076e+00, -7.82473445e-01,  1.39130369e-01, -1.88474381e+00,
     -1.57926226e+00,  6.76095903e-01,  7.99891353e-01,  2.52676696e-01,
      7.71672785e-01,  1.29661608e+00,  2.12828502e-01, -8.35336447e-01,
      2.18707371e+00, -1.00460708e+00, -2.50636518e-01, -7.12184548e-01,
      1.23993087e+00, -1.55890131e+00, -6.44942045e-01,  1.16504826e-01,
      2.53171563e+00,  2.20096874e+00,  5.50863221e-02, -6.24714613e-01,
     -1.16663885e+00,  8.13338498e-04, -7.39723444e-01, -6.57191753e-01,
     -3.26518029e-01,  4.04749781e-01,  1.29682720e+00,  1.70414293e+00],
dtype=float32)
(96,)
```

Figure 2. Word vector representation

The POS tagging is correctly identified for each word in the sentence. It is useful to understand accuracy which relates to input sentence and construct output responses.

Predicted RELATION

The relations are predicted from the identified Noun words with the interaction keyword. Some of the interaction or relation keywords are associates, integrates, inhibits, activates, etc. Example 2 shows the prediction relation using the spacy model. In table 2shows the extraction of relations with different patterns.

Example 2:

Sentence	Pattern	Predicted relations
CHD8 activates BRG1 associated SWI/SNF activate CHD7	[{'POS':'PROPN'}, {'LOWER':'associated'}, {'POS':'PROPN'}]	BRG1 associated SWI

V. CONCLUSION

The study focuses on the relation extraction from sentences using spacy package. The spacy is a reliable NLP framework that quickly became the standard for doing NLP in Python. This package increases the speed, accuracy and extensibility. The framework consists of following phases such as Data modeling, Loading Libraries, preprocessing the sentences, Entity detection and relation extraction phase. The relations are identified and extracted based on the NOUN words and patterns using the defined pattern. The pattern includes the following fields such as POS tagging, interaction keyword and POS tagging. This approach identifies only the direct relations between noun words; otherwise it does not find. The method tested with many interaction keywords as follows: “associated”, “activated”, “activates”, “reveal”, “induce”, “inhibition”, “promotes”, “revealed”, “inhibits”, “suggest”, “relieves”, “ associated protein”, “exhibits”, “stimulates”, “associated with”, “inhibition promotes”, “reduced”, blocked”, “associated with”. The interaction keyword maybe single word or combination of two or three words. All the interaction keywords are NOUN words. The dataset contains totally 1000 sentence; trained with 900 sentences and tested with 100 sentences. The results are verified and validated with the benchmarking databases. From this analysis, the method achieves higher (best) accuracy when compared to other approaches. In future, the hidden relationships are extracted using other deep learning related package.

REFERENCES

1. Prateek
2. Joshi, “How search engines like Google Retrieve Results: Introduction to Information Extraction using Python and Spacy”, September 2019.
3. Tutorial: Text Classification in python using spacy
4. Bhasuran, B., Subramanian, D., & Natarajan, J. (2018). Text mining and network analysis to fi nd functional associations of genes in high altitude diseases. Computational Biology and Chemistry, 75, 101–110. https://doi.org/10.1016/j.compbiolchem.2018.05.002
5. Cui, H., Zhang, M., Yang, Q., Li, X., Liebman, M., Yu, Y., &Xie, L. (2018). The Prediction of Drug-Disease Correlation Based on Gene Expression Data. 2018.
6. Gupta, S., Dingerdissen, H., Ross, K. E., Hu, Y., Wu, C. H., &Mazumder, R. (2018). Original article DEXTER : Disease-Expression Relation Extraction from Text. 1–17. https://doi.org/10.1093/database/bay045
7. Chang, Y., Chu, C., Su, Y., & Chen, C. C. (2016). Original article PIPE : a protein – protein interaction passage extraction module for BioCreative challenge. 1–16. https://doi.org/10.1093/database/baw101
8. Bravo, À., Piñero, J., Queralt-rosinach, N., Rautschka, M., & Furlong, L. I. (2015). Extraction of relations between genes and diseases from text and large-scale data analysis : implications for translational research. 1–17. https://doi.org/10.1186/s12859015-0472-9
9. Bchir, A. (2015). Extraction of Drug-disease relations from MEDLINE Abstracts. (October). https://doi.org/10.1109/WCCIT.2013.6618759
10. Xu, R., & Wang, Q. (2013). Large-scale extraction of accurate drug-disease treatment pairs from biomedical literature for drug repurposing.
11. Kim, Jeongkyun, So, S., Lee, H., Park, J. C., Kim, J., & Lee, H. (2013). DigSee : disease gene search engine with evidence sentences (version cancer). 41(June), 510–517. https://doi.org/10.1093/nar/gkt531

12. Chun, H. W., Tsuruoka, Y., Kim, J. D., Shiba, R., Ata, N. N., Hishiki, T., &Tsujii, J. (2006). Extraction of gene-disease relations from medline using domain dictionaries and machine learning. Proceedings of the Pacific Symposium on Biocomputing 2006, PSB 2006, 15, 4–15. https://doi.org/10.1142/9789812701626_0002
13. Complete Guide to spacy

AUTHORS PROFILE

G Suganya, received gold madal for her Master degree Computer Science and pursuing PhD in Bharathiar University. Her research interests include Information Extraction, Information Retrieval and Bioinformatics.

Dr. R. Porkodi, received MCA degree and pursued Ph.D in Bharathiar University. She received UGC grant for her research study. She is the member of many academic bodies. She is the life member in computer society in India, member in IAENG and IACSIT. She acted as a committee member/ resource person / coordinator for various research conferences / events / Faculty development programmes. She published many articles in various reputed journals. Her research interests include Data mining, NLP, Image mining, CBMIR, Hyperspectral remote sensing and Bioinformatics.

Table 2 Predicting the relations

Sentence	After Pre-Processing	Pattern	Meaning	POS tagging	Predicted relation
BRG1 associated SWI/SNF complexes that in turn activate CHD7	BRG1 associated SWI/SNF complexes turn activate CHD7	Pattern1= [{'POS':'NOUN'} , {'LOWER': 'associated'}, {'POS': 'PROPN'}]	It extracts the relation which contains the keyword as "associated" between NOUN and PROPN	BRG1 -->nsubj --> PROPN associated --> ROOT --> VERB SWI -->nmod --> PROPN / -->punct --> SYM SNF --> compound --> PROPN complexes -->dobj --> NOUN turn -->pobj --> NOUN activate -->relcl --> VERB CHD7 -->dobj --> PROPN	BRG1 associated SWI
Western blot results of autophagy-associated protein (LC3 II, Beclin-1) and apoptosis associated proteins caspase3, Bcl-2 revealed AG-1031 could activate apoptotic signal pathway via inhibiting autophagy process in cancer cells	Western blot results autophagy-associated protein (LC3 II, Beclin-1) apoptosis associated proteins caspase3, Bcl-2 revealed AG-1031 activate apoptotic signal pathway via inhibiting autophagy process cancer cells	Pattern1= [{'POS':'PROPN' }, {'LOWER': 'reveal'}, {'POS': 'PROPN'}]	It extracts the relation which contains the keyword as "reveal" between NOUN and PROPN	Western -->amod --> ADJ blot --> compound --> NOUN results -->nsubj --> NOUN autophagy -->npadvmod --> ADJ - -->punct --> PUNCT associated -->amod --> VERB protein -->pobj --> NOUN (-->punct --> PUNCT LC3 --> compound --> PROPN II -->appos --> PROPN , -->punct --> PUNCT Beclin-1 -->appos --> PROPN) -->punct --> PUNCT apoptosis -->conj --> NOUN associated -->amod --> VERB proteins -->dobj --> NOUN (-->punct --> PUNCT caspase3 -->appos --> PROPN , -->punct --> PUNCT Bcl-2 -->nmod --> PROPN revealed --> ROOT --> VERB AG-1031 -->nsubj --> PROPN activate -->ccomp --> VERB apoptotic -->amod --> ADJ signal --> compound --> ADJ pathway -->dobj --> NOUN via --> prep --> ADP inhibiting -->pcomp --> VERB autophagy --> compound --> ADJ process -->dobj --> NOUN cancer --> compound --> NOUN cells -->pobj --> NOUN	Bcl-2 revealed AG-1031
SD can increase active oxygen species and as a result, damage mitochondria induce apoptosis	SD increase active oxygen species result, damage mitochondria induce apoptosis	Pattern1= [{'POS':'NOUN'} , {'LOWER': 'induce'}, {'POS': 'NOUN'}]	It extracts the relation which contains the keyword as "induce" between NOUN and NOUN	SD -->nsubj --> NOUN increase --> ROOT --> VERB active -->amod --> ADJ oxygen --> compound --> NOUN species -->dobj --> NOUN result -->pobj --> NOUN , -->punct --> PUNCT damage --> compound --> NOUN mitochondria-->npadvmod-->NOUN induce -->conj --> VERB apoptosis -->dobj --> NOUN	mitochondria induce apoptosis

Furthermore, CaMKII is also a tau kinase and its dysregulation associated with Alzheimer progression and CaMKII inhibits tau-microtubule interaction by tau phosphorylation	Furthermore, CaMKII tau kinase dysregulation associated with Alzheimer progression CaMKII inhibits tau-microtubule interaction tau phosphorylation	Pattern1= [{'POS':'NOUN'} , {'LOWER': 'associated'}, {'LOWER':'with' }, {'POS': 'PROPN'}]	It extracts the relation which contains the keyword as "associated with" between NOUN and PROPN	Furthermore -->advmod --> ADV , -->punct --> PUNCT CaMKII -->nsubj --> NOUN tau --> compound --> NOUN kinase -->attr --> NOUN dysregulation -->conj --> NOUN associated -->conj --> VERB with --> prep --> ADP Alzheimer --> compound --> PROPN progression -->pobj --> NOUN CaMKII -->conj --> PROPN inhibits -->conj --> VERB tau --> compound --> PROPN - -->punct --> PUNCT microtubule --> compound --> ADJ interaction -->dobj --> NOUN tau --> compound --> PROPN phosphorylation -->pobj --> NOUN	dysregulation associated with Alzheimer
Co-administration with selective cannabinoidreceptor revealed PrNMI antiallodynic effects are mediated by CB1 receptor (CB1R) activation	Co-administration with selective cannabinoidreceptor revealed PrNMI antiallodynic effects mediated CB1 receptor (CB1R) activation	Pattern1= [{'POS':'NOUN'} , {'LOWER': 'revealed'}, {'POS': 'PROPN'}]	It extracts the relation which contains the keyword as "revealed" between NOUN and PROPN	Co --> dep --> PROPN - --> dep --> NOUN administration -->nsubj --> NOUN with --> prep --> ADP selective -->amod --> ADJ cannabinoidreceptor-->compound-->NOUN revealed --> ROOT --> VERB PrNMI -->poss --> PROPN antiallodynic -->amod --> ADJ effects -->nsubjpass --> NOUN mediated -->ccomp --> VERB CB1 --> compound --> PROPN receptor -->pobj --> NOUN (-->punct --> PUNCT CB1R -->nmod --> PROPN) -->punct --> PUNCT activation -->appos --> NOUN	cannabinoidreceptor revealed PrNMI
These findings suggest that PTEN inhibition promotes angiogenesis in HUVECs after exposure to OGD and this enhancing effect might be achieved through activation of the Akt signal cascade	These findings suggest PTEN inhibition promotes angiogenesis HUVECs exposure OGD enhancing effect might achieved through activation Akt signal cascade	Pattern1= [{'POS':'PROPN' }, {'LOWER': 'inhibition'}, {'LOWER': 'promotes'} {'POS': 'NOUN'}]	It extracts the relation which contains the keyword as "inhibition promotes" between PROPN and NOUN	These --> det --> DET findings -->nsubj --> NOUN suggest --> ROOT --> VERB PTEN --> compound --> PROPN inhibition -->nsubj --> VERB promotes -->ccomp --> VERB angiogenesis -->dobj --> NOUN HUVECs -->pobj --> NOUN exposure -->pobj --> NOUN OGD -->pobj --> PROPN enhancing -->amod --> VERB effect -->nsubjpass --> NOUN might --> aux --> VERB achieved -->ccomp --> VERB through --> prep --> ADP activation -->pobj --> NOUN Akt --> compound --> PROPN signal --> compound --> NOUN cascade -->pobj --> NOUN	PTEN inhibition promotes angiogenesis

Mechanistically, BMX bypasses the suppressor of SOCS3 inhibition JAK2, whereas NPCs dampen the JAK2-mediated STAT3 activation via the negative regulation by SOCS3, providing a molecular basis for targeting BMX by ibrutinib to specifically eliminate GSCs while preserving NPCs	Mechanistically, BMX bypasses suppressor SOCS3 inhibition JAK2, whereas NPCs dampen JAK2-mediated STAT3 activation negative regulation SOCS3, providing molecular basis targeting BMX ibrutinib specifically eliminate GSCs while preserving NPCs	Pattern1= [{'POS':'NOUN'} , {'LOWER': 'inhibition'}, {'POS': 'PROPN'}]	It extracts the relation which contains the keyword as "inhibition" between NOUN and PROPN	Mechanistically-->advmod --> ADV , -->punct --> PUNCT BMX -->nsubj --> PROPN bypasses --> ROOT --> VERB suppressor -->dobj --> NOUN SOCS3 -->nsubj --> NOUN inhibition -->dobj --> NOUN JAK2 -->appos --> PROPN , -->punct --> PUNCT whereas --> mark --> SCONJ NPCs -->nsubj --> NOUN dampen -->advcl --> VERB JAK2-mediated -->dobj --> PROPN STAT3 --> compound --> PROPN activation --> ROOT --> NOUN negative -->amod --> ADJ regulation -->pobj --> NOUN SOCS3 -->pobj --> PROPN , -->punct --> PUNCT providing -->advcl --> VERB molecular -->amod --> ADJ basis -->dobj --> NOUN targeting -->pcomp --> VERB BMX -->dobj --> PROPN ibrutinib -->pobj --> NOUN specifically -->advmod --> ADV eliminate -->advcl --> VERB GSCs -->dobj --> NOUN while --> mark --> SCONJ preserving -->advcl --> VERB NPCs -->dobj --> NOUN	SOCS3 inhibition JAK2
Western blot results of autophagy associated protein LC3 II, Beclin-1 and apoptosis-associated proteins	Western blot results autophagy associated protein LC3 II, Beclin-1 apoptosis-associated proteins	Pattern1= [{'POS':'PROPN' }, {'LOWER': 'associated'}, {'LOWER':'protein'}, {'POS': 'PROPN'}]	It extracts the relation which contains the keyword as "associated protein" between PROPN words	Western -->amod --> ADJ blot --> compound --> NOUN results -->nsubj --> NOUN autophag--> compound --> PROPN associated -->amod --> VERB protein --> compound --> NOUN LC3 --> compound --> PROPN II -->pobj --> PROPN , -->punct --> PUNCT Beclin-1 -->conj --> PROPN apoptosis -->npadvmod --> NOUN - -->punct --> PUNCT associated -->amod --> VERB proteins -->conj --> NOUN	autophagy associated protein LC3
Cellular MYPOP relieves HPV16 infection, demonstrating that MYPOP acts as a restriction factor	Cellular MYPOP relieves HPV16 infection, demonstrating MYPOP acts restriction factor	Pattern1= [{'POS':'PROPN' }, {'LOWER': 'relieves'}, {'POS': 'NOUN'}]	It extracts the relation which contains the keyword as "relieves" between PROPN and NOUN	Cellular -->amod --> ADJ MYPOP --> compound --> PROPN relieves --> ROOT --> VERB HPV16 --> compound --> NOUN infection -->pobj --> NOUN , -->punct --> PUNCT demonstrating -->advcl --> VERB MYPOP -->nsubj --> PROPN acts -->ccomp --> VERB restriction --> compound --> NOUN factor -->pobj --> NOUN	MYPOP relieves HPV16
Moreover, overexpression of MYPOP blocks colony formation of HPV and non-virally transformed keratinocytes, suggesting that MYPOP exhibits tumor suppressor properties	Moreover, overexpression MYPOP blocks colony formation HPV non-virally transformed keratinocytes, suggesting MYPOP exhibits tumor suppressor properties	Pattern1= [{'POS':'PROPN' }, {'LOWER': 'exhibits'}, {'POS': 'NOUN'}]	It extracts the relation which contains the keyword as "exhibits" between PROPN and NOUN	Moreover -->advmod --> ADV , -->punct --> PUNCT overexpression -->nsubj --> NOUN MYPOP --> compound --> PROPN blocks -->pobj --> VERB colony --> ROOT --> NOUN formation -->dobj --> NOUN HPV -->pobj --> PROPN non --> dep --> ADJ - --> dep --> ADJ virally -->advmod --> ADV transformed -->amod --> VERB keratinocytes -->conj --> NOUN , -->punct --> PUNCT suggesting -->advcl --> VERB MYPOP -->nsubj --> PROPN exhibits -->ccomp --> VERB tumor --> compound --> NOUN suppressor --> compound --> NOUN properties -->dobj --> NOUN	MYPOP exhibits tumor
Our findings suggest that E7 stimulates MYPOP degradation	Our findings suggest E7 stimulates MYPOP degradation	Pattern1= [{'POS':'PROPN' }, {'LOWER': 'stimulates'}, {'POS': 'PROPN'}]	It extracts the relation which contains the keyword as "stimulates" between PROPN words	Our -->poss --> DET findings -->nsubj --> NOUN suggest --> ROOT --> VERB E7 -->nsubj --> PROPN stimulates -->ccomp --> VERB MYPOP --> compound --> PROPN degradation -->dobj --> NOUN	E7 stimulates MYPOP

skincancer associated with DLE	skincancer associated with DLE	Pattern1= [{'POS':'NOUN'} , {'LOWER': 'associated'}, {'LOWER':'with' }, {'POS': 'PROPN'}]	It extracts the relation which contains the keyword as "associated with" between NOUN and PROPN	skincancer --> compound --> NOUN associated -->acl --> VERB with --> prep --> ADP DLE -->appos --> PROPN	skincancer associated with DLE
Therefore, discoid lupus inflammation promotes skincancer in high-risk DLE patients	Therefore, discoid lupus inflammation promotes skincancer high-risk DLE patients	Pattern1= [{'POS':'NOUN'} , {'LOWER': 'promotes'}, {'POS': 'NOUN'}]	It extracts the relation which contains the keyword as "promotes" between NOUN words	Therefore -->advmod --> ADV , -->punct --> PUNCT discoid -->npadvmod --> ADJ lupus --> compound --> NOUN inflammation -->nsubj --> NOUN promotes -->nsubj --> VERB skincancer --> compound --> NOUN high -->amod --> ADJ - -->punct --> PUNCT risk --> compound --> NOUN DLE --> compound --> PROPN patients -->pobj --> NOUN	inflammation promotes skincancer
Importantly tacrolimus reduced inflammation and blocked cancer development in MRL/lpr mice	Importantly tacrolimus reduced inflammation blocked cancer development MRL/lpr mice	Pattern1= [{'POS':'NOUN'} , {'LOWER': 'reduced'}, {'POS': 'NOUN'}]	It extracts the relation which contains the keyword as "reduced" between NOUN words	Importantly -->advmod --> ADV tacrolimus --> compound --> NOUN reduced --> ROOT --> VERB inflammation -->dobj --> NOUN blocked -->amod --> VERB cancer --> compound --> NOUN development -->conj --> NOUN MRL -->nmod --> PROPN / -->punct --> SYM lpr --> compound --> PROPN mice -->pobj --> NOUN	tacrolimus reduced inflammation