

# Discovering New Knowledge on the Neuroendocrine System by Extracting Knowledge from PubMed using Natural Language Processing

## 1 Abstract

This section will briefly summarize:

- The need for using AI/NLP to gather biomedical knowledge, using neuroendocrine system as an example.
- The methods the group used to gather, process, and visualize knowledge.
- The impact of using these methods, and how they can be improved in the future.

## 2 Introduction and Background

- Discuss the need to use software to help researchers to see the progress made in knowledge about the neuroendocrine system.
- This is essentially a “Big Data” problem.
  - **Volume:** Large number of abstracts on PubMed
  - **Velocity:** New articles on aging published daily
  - **Variety:** Text, illustrations, images, etc. (although we only deal with textual data)
- Briefly summarize how the group approaches the problem.

## 3 Knowledge Extraction and NLP

- How we use keywords to filter and download abstracts, streaming in abstracts as they are published (RSS/Feedly?).
- How we use Stanford NLP library to parse sentences.
- How we perform pattern matching on the parse tree in Scala to derive relations.

## 4 Relation Processing

- How we can compile relations and export them to CSV format.
- How we filter relations, or triples, using a limited set of predicates and subject/objects that are only cells and substances.
- How we can relate relations to the body anatomy tree.

## 5 Relation Visualization

- Data representation using nodes and edges, and how clustering can be applied.
- Visualization using Gephi.
- Description of web app features.
- Discuss interesting results we found through the data visualization.

## **6 Conclusion**

- Describe what we learned about knowledge of neuroendocrine system as a result of the methods we used.
- How the methods we used help us to accomplish our task, and the possibility of generalizing it to help apply it to related studies.
- Describe how we can improve the process.

## **7 Bibliography**

## **8 Appendices**