

Artificial Intelligence And Open Science In Research Software Engineering

Deliverable 2

Sebastian Bayona
Alejandro Ayuso
Daniel Cabrera
Alejandro Morán

<https://github.com/dan415/openscience-grupo>

Papers selection

There are papers from 2 distinct research fields, and the main reason for selecting them is their complete metadata: Acknowledgments, authors, affiliations, etc.

arXiv:2305.06541v1 [cs.LG] 11 May 2023

Spectral Clustering on Large Datasets: When Does it Work? Theory from Continuous Clustering and Density Cheeger-Buser

Timothy Chu

Gary L. Miller

Noel J. Walkington

Biomedical Signal Processing and Control 47 (2019) 312–323



Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

journal homepage: www.elsevier.com/locate/bspc



Speech emotion recognition using deep 1D & 2D CNN LSTM networks

Jianfeng Zhao^{a,b}, Xia

^a School of Electronics and Information
^b School of Information Engineering, F

Circulation

Volume 92, Issue 5, 1 September 1995, Pages 1084–1088
<https://doi.org/10.1161/01.CIR.92.5.1084>



ARTICLE INFO

Article history:
Received 12 July 2017
Received in revised form 26 July 2018
Accepted 27 August 2018
Available online 11 September 2018

Keywords:
Speech emotion recognition
CNN LSTM network
Raw audio clips
Log-mel spectrograms

ARTICLE

Infiltrates of Activated Mast Cells at the Site of Coronary Atheromatous Erosion or Rupture in Myocardial Infarction

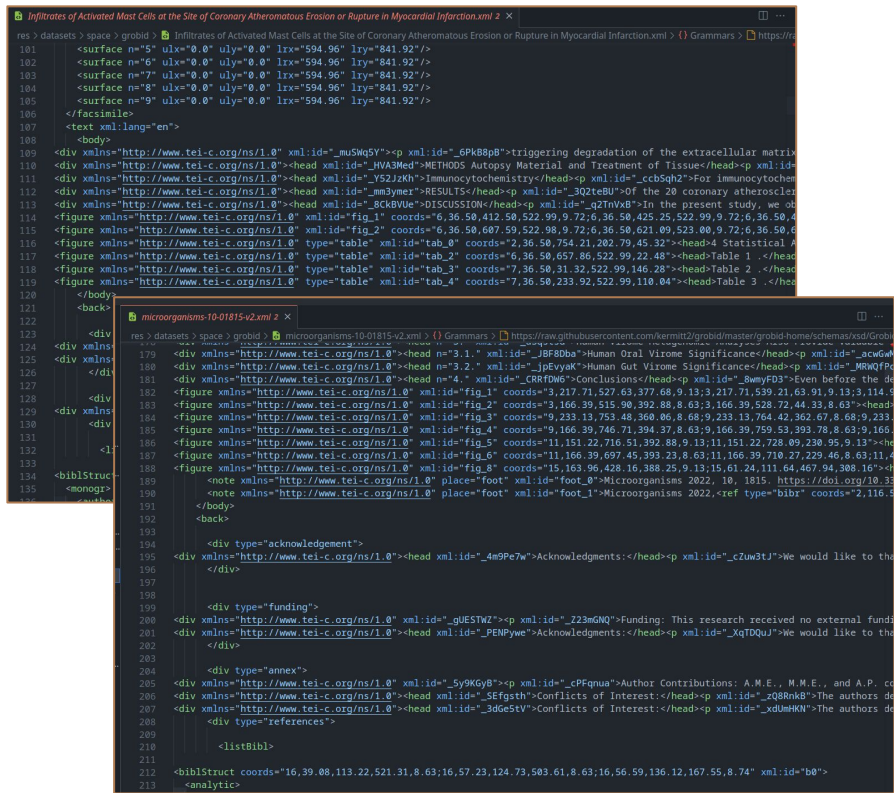
Petri T. Kovanen, Maija Kaartinen, and Timo Paavonen

ABSTRACT: Background Erosion and rupture of coronary atheromas are the events preceding the vast majority of acute coronary syndromes. The shoulder regions of atheromas, the sites at which erosion or rupture is most likely to occur, are the sites at which mast cells accumulate. These cells are filled with neutral proteases capable of triggering extracellular matrix degradation via activation of matrix metalloproteinases. To obtain more direct evidence for the participation of mast cells in the acute coronary syndromes, we quantified the numbers of mast cells at eroded or ruptured sites of coronary atheromas in patients who died of myocardial infarction. **Methods and Results** In specimens of coronary arteries from 20 patients who had died of acute myocardial infarction, the site of atheromatous erosion or rupture was identified. The specimens were stained with monoclonal antibodies against the two major proteases of mast cells, tryptase and chymase, and against macrophages, T lymphocytes, and smooth muscle cells. At the immediate site of erosion or rupture, mast cells amounted to 6% of all nucleated cells, in the adjacent atheromatous area to 1%, and in the unaffected intimal area to 0.1%. The proportions of these mast cells that were activated, ie, had been stimulated to degranulate and release some of their tryptase and chymase contents, were 86% at the site of erosion or rupture, 63% in the adjacent atheromatous area, and 27% in the unaffected intima. At the site of erosion or rupture, the numbers of macrophages and T lymphocytes were also increased, but the number of smooth muscle cells was decreased. **Conclusions** The accumulation of activated mast cells (200-fold more than in the unaffected coronary intima) at the site of atheromatous erosion or rupture suggests that in thrombotic coronary occlusion the role played by mast cells is significant.

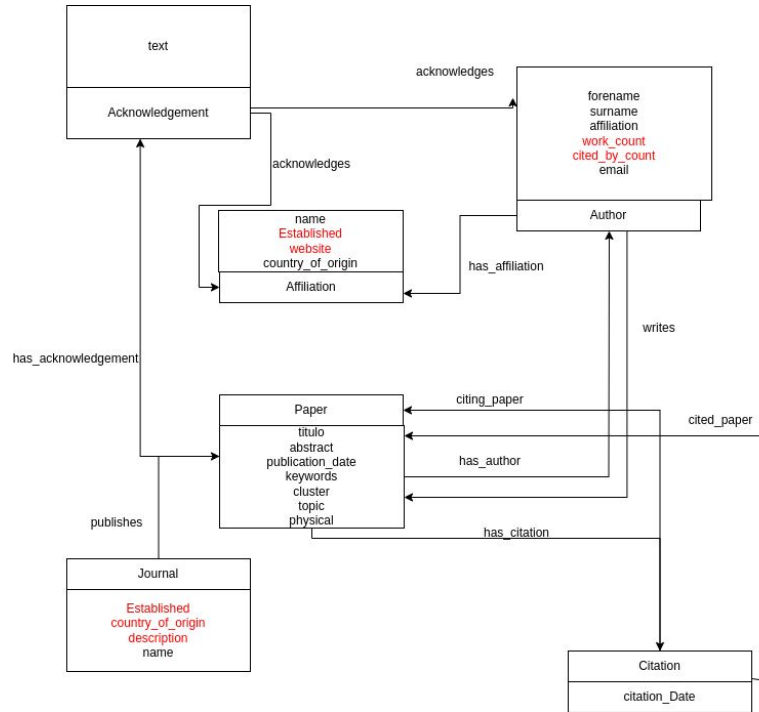
Key Words: atherosclerosis ■ chymase ■ tryptase ■ mast cells ■ atherosclerosis ■ myocardial infarction

Leveraging GROBID for Knowledge Graph Construction: Extracting Relevant Information from Selected Papers

In the realm of technical and scientific publications, extracting structured information from raw documents like PDFs is crucial for various research and knowledge management tasks. GROBID, a powerful machine learning library, enables the extraction, parsing, and re-structuring of PDFs into structured XML/TEI encoded documents. This presentation focuses on utilizing GROBID to extract relevant information from selected papers and leveraging this data to construct a knowledge graph. By employing GROBID's functionalities.



Ontology



Wikidata



Affiliation (Established, Website)

Journal (Established, Country_of_origin, Description)



OpenAlex



Author (Cited_by_count, Works_count)



OpenAlex

Research Knowledge Graph

1. Get classes and properties from the xml already processed
2. Paper Space for a more useful representation
3. Obtain Knowledge Graph from Paper Space →

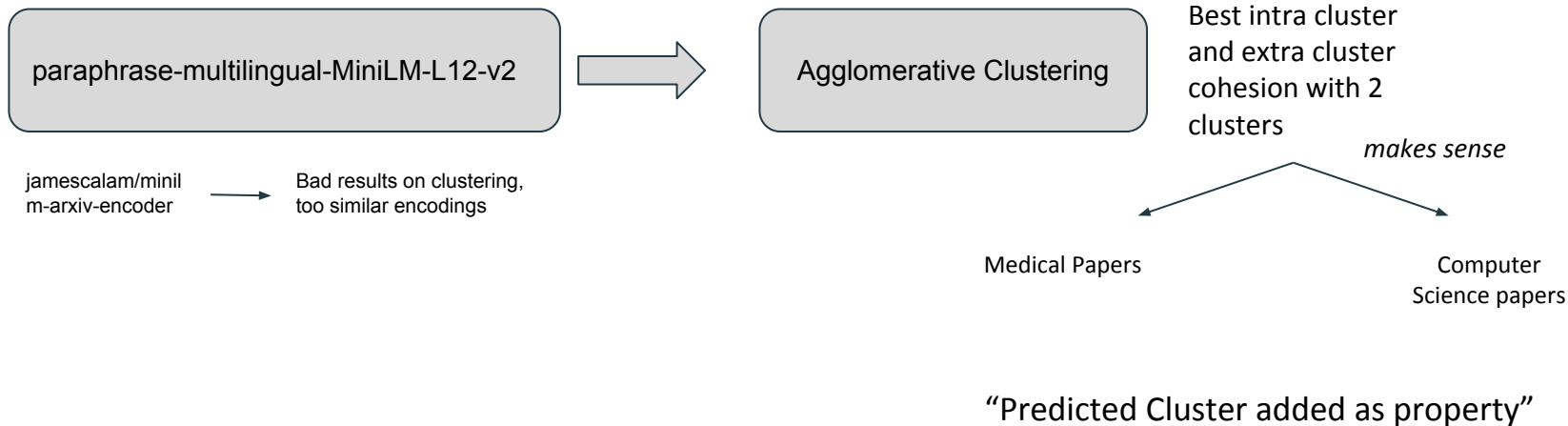
JSON-LD

XML processed by Grobid

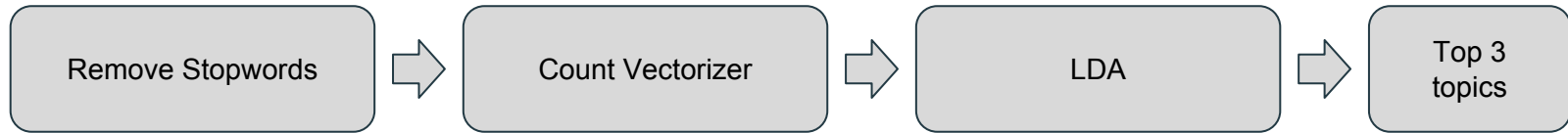
Representation to operate with

RDFParser -> JSON-LD

Clustering

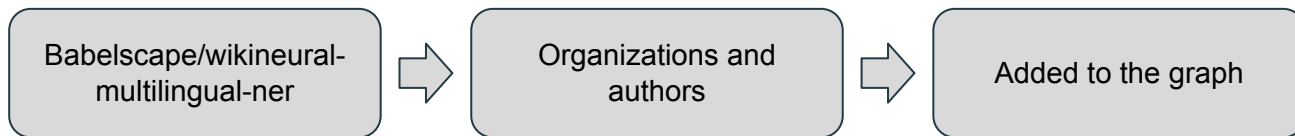


Topic Modeling



“Topics added as property”

Entity Recognition



Experiment as a Research Object: using ya2ro

```
project.yaml > {} participants > {} 2 > web
1 # Mandatory field
2 type: "project"
3
4 title: "Project - Artificial Intelligence And Open Science In Research Software Engineering Deliverable 2"
5
6 goal: "Advanced data analysis on research publications: Given a corpus of 30 papers, group them according to common
7
8 social_motivation: "In the field of healthcare and medical research, clinical trials play a crucial role in advancing
9
10 #sketch: "images/sketch_ya2ro.jpg"
11
12 areas:
13   - "Area 1: Use information on clinical trials associated with the publications to identify trends in clinical res
14
15 activities:
16   - "Subtask 1: Papers selection. There are papers from 2 distinct research fields, and the main reason for selecti
17   - "Subtask 2: Leveraging GROBID for Knowledge Graph Construction: Extracting Relevant Information from Selected P
18   - "Subtask 3: Ontology. Diagram of the ontology that was defined for the project"
19   - "Subtask 4: Transform the classes and properties from the xml already processed (by GROBID) to the \Paper Space
20   - "Subtask 5: Enriching with wikidata (affiliation, website and Journal) and OpenAlex (author's count of citation
21   - "Subtask 6: Obtain Knowledge Graph from Paper Space -> JSON-LD"
22   - "Subtask 7: Clustering"
23   - "Subtask 8: Topic Modeling"
24   - "Subtask 9: Entity Recognition"
25
26 demo:
27   - name: "Git Repository"
28     link: https://github.com/dan415/openscience-grupo
29     description: "Find the docs and demos in the github repository."
30
31 datasets:
32   - link: https://drive.google.com/drive/folders/1ACPAXCJiIzueSEnA3p1kjZoMwzqGV4qG7usp#share\_link
33     name: "Dataset 1"
34     description: "List of selected papers"
35
36 software:
37   - link: https://github.com/dan415/openscience-grupo
38     name: "openscience-grupo"
39     description: "Main code repository"
40     license: "MIT-license"
41
```



Project - Artificial Intelligence And Open Science In Research Software Engineering Deliverable 2

Goal

Advanced data analysis on research publications: Given a corpus of 30 papers, group them according to common themes, link them in a Research Knowledge Graph (RKG) together with their metadata and funding information.

Social Motivation

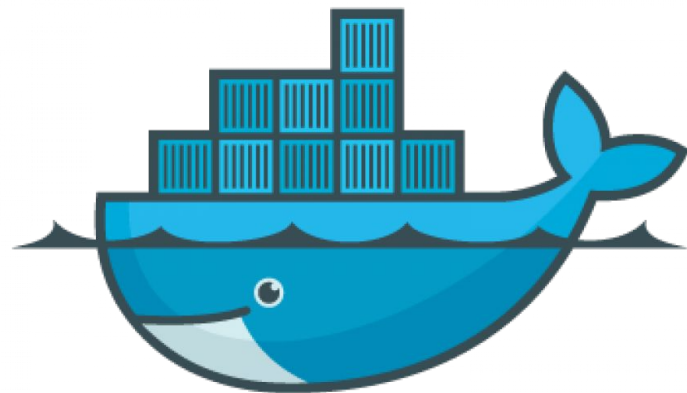
In the field of healthcare and medical research, clinical trials play a crucial role in advancing our understanding of diseases, testing new treatments, and improving patient outcomes. As the volume of clinical research continues to grow, it becomes essential to identify trends and gain valuable insights from these studies.

```
output > project > {} ro-crate-metadata.json > ...
1 {
2   "@context": "https://w3id.org/ro/crate/1.1/context",
3   "@graph": [
4     {
5       "@id": "ro-crate-metadata.json",
6       "@type": "CreativeWork",
7       "about": {
8         "@id": ". /",
9       },
10      "conformsTo": {
11        "@id": "https://w3id.org/ro/crate/1.1"
12      },
13      "description": "RO-Crate Metadata File Descriptor (this file)"
14    },
15    {
16      "@id": ". /",
17      "@type": "Dataset",
18      "dataType": "Project",
19      "author": [
20        {
21          "@id": "#daniel_cabrera"
22        },
23        {
24          "@id": "#alejandro_ayuso"
25        },
26        {
27          "@id": "#alejandro_morales"
28        },
29        {
30          "@id": "#sebastian_bayona"
31        }
32      ],
33      "description": "Advanced data analysis on research publications: Given a corpus of 30 papers, group them according to common themes, link them in a Research Knowledge Graph (RKG) together with their metadata and funding information."
34      "hasPart": [
35        {
36          "@id": "#openscience-grupo"
37        },
38      ]
39    }
40  ]
41 }
```

ro-crate-metadata.json

Best practices

- Description
- Registries
- Releases
- Requiriments
- Installation
- Usage
- Tests



docker



Demo