

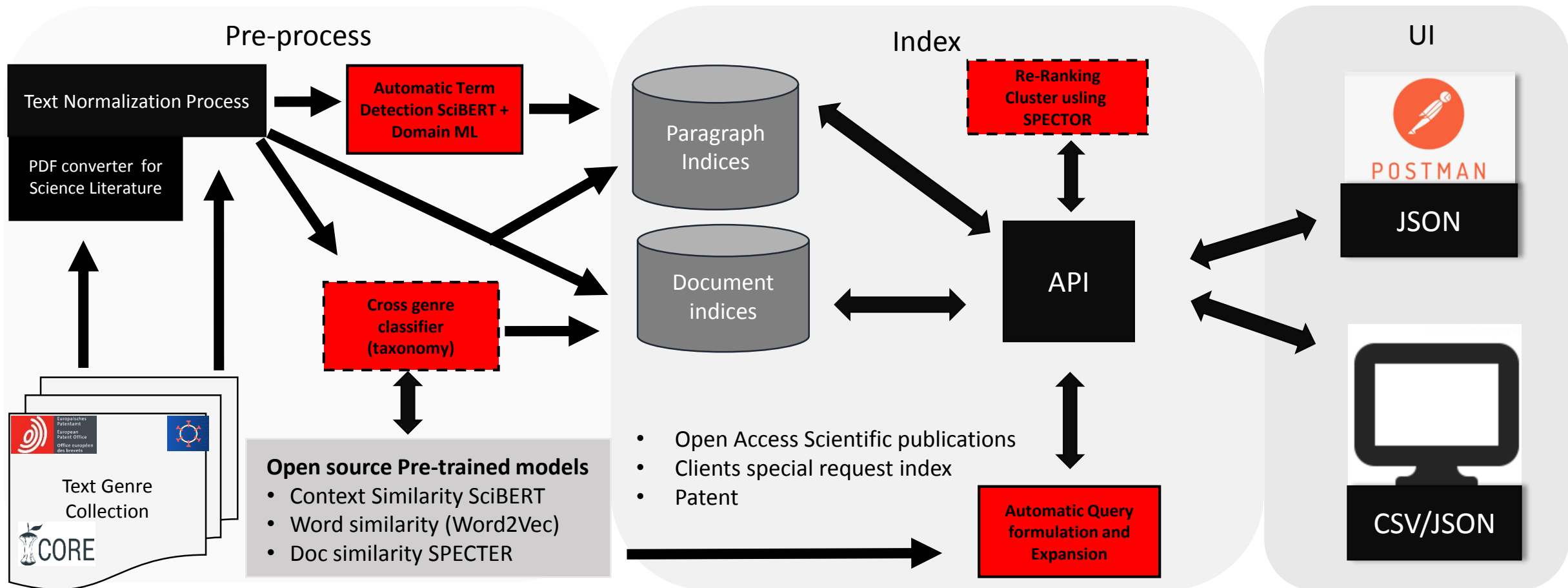
Artificial Researcher

Outline

- Artificial Research Passage Retrieval Services
 - Technology & Usage
- Artificial Researcher Ontology Services
 - Technology & Usage
- Artificial Researcher NLP-toolkit Service
 - Release and rest API access

Artificial Researcher Passage Retrieval Service

Artificial Researcher Passage Retrieval Services



Open source

Software

Language

Open source container

GROBID Documentation

Ai2

GitHub: allnasa/sciBERT: A. github.com

Solr

Lucene™

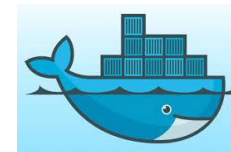
python

GO

Java

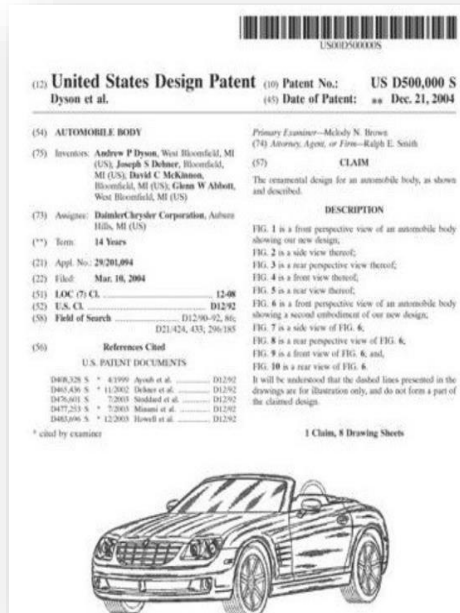
Perl

C++



kubernetes

Automatic Query Formulation and Expansion



Example of automatic query generation

<QUERY>
 (conured OR clutch OR connectability OR nmofs OR fclp OR dnsr OR slippage OR anda OR rotational OR acceleration OR backlash OR subordinate OR estimating OR ure OR brake OR torque OR stopped OR vehicle OR wheel OR command OR outputting OR estimate OR check OR ner OR driving OR pedal OR wheels OR shaft OR prohibiting OR determining OR estimated OR prescribed OR stopping OR elapse OR output OR controller OR rotating OR acceleration AND
 ("vehicle driving force control apparatus" OR "dr OR "clutch connection command" OR "rear wheel path" OR "output rotational speed" OR "input ro "detected parameter" OR "generation load torqu "determination occurrence" OR "four-wheel driv proceed" OR "wheel speed sensor" OR "output s "backlash elimination" OR "drive mode switch" (" range" OR "transition time" OR "wheel speed" C connection" OR "motor torque" OR "generator le "high rate" OR "electric motor" OR "throttle ope force" OR "connected state" OR "previous equat "prescribed rotational speed difference" OR "12 "disconnected state" OR "electric clutch" OR "fo
 </QUERY>

Automatic query expansion terms

brake pedal:
 vehicle operating pedal,
 conventional hydraulic brake system
 pedal devices
 position brake actuating member
 brake actuating member
 hydraulically-assisted rack pinion steering gear
 brake operating member
 conventional braking system
 pair pedals

accelerator pedal
 case pedal device
 pedal device

Claims (1)

1. The ornamental design for an automobile body, as shown and described.

Passage Retrieval Service

<https://passageretrieval.artificialresearcher.com/>

- Standard Index Collections
 - Open Access scientific articles
 - access to 204,582,649 – CORE.uk
 - Patents
 - EP full-text data for text analytics - European Patent Office
- Indexing domain-specific collections
 - e.g. COVID-19 Open Research Dataset (Wang et al 2019, Allen Institute for AI)

Artificial Researcher

Welcome to Artificial Researcher's Passage Retrieval Service

On this platform we provide a demo version of our unique passage retrieval service. Our service uses innovative text mining technologies based on 15 years of collected know-how and research that gives you state of the art machine learning. The data collections available to you in this demo are: COVID-19 Open Research Dataset (CORD-19) data set, a sample set from the EP Full-Text Data for Text Analytics, and a sample composed of different scientific publications within technical and medical science provided by CORE.uk (Science). [\(White Paper\)](#). Please give us feedback and let us know what you think to help us improve and develop a service you will love to use. For access our larger collections, which we provide via APIs, please contact demo@artificialresearcher.com

Artificial Researcher – IT GmbH

Select collection*

Meta data

Name	Affiliation	Year
<input type="text" value="Miller OR Zhang"/>	<input type="text" value="Johns Hopkins University School of Medicine"/>	<input type="text" value="2020"/>

Taxonomy

Keywords

Title

Text*

Boolean search (To learn more about boolean search read our [White Paper](#))

Passage Retrieval Service

- Direct access to
 - Relevant paragraphs
 - Abstracts
 - Bibliographic information (author, year, affiliation)
 - Links to full text documents
- Possibility to select and download paragraphs including all bibliographic data, **including the identified technical terms**, subject classification terms, scientific classifications
- *The sample collections on the showcase page are updated every 4th day*

The screenshot displays the Artificial Researcher interface with search results for the query 'Commentary Avian Influenza Virchow's Reminder'. The interface includes a search bar at the top with 'New search', 'Edit query', and 'Feedback' buttons. Below the search bar, there are three search results, each with a title, a brief abstract, keywords, a link to the full text, and a 'Select' button. The first result is titled 'Commentary Avian Influenza Virchow's Reminder' and discusses the collaboration of scientists and regulatory entities. The second result is also titled 'Commentary Avian Influenza Virchow's Reminder' and discusses the zoonosis of avian influenza. The third result is titled 'Commentary Avian Influenza Virchow's Reminder' and discusses the interface of human and animal health. Each result includes a 'Select' button and a 'Report data error' button.

Request via Postman

<http://swagger.artificialresearcher.com/>

GET <https://passageretrieval-api.artificialresearcher.com/search?domain=patent&simpleAuth=<api key>&userI> Send

Params ● Authorization Headers (6) Body Pre-request Script Tests Settings

<input checked="" type="checkbox"/>	domain	patent
<input checked="" type="checkbox"/>	simpleAuth	<api key>
<input checked="" type="checkbox"/>	userDefined	true
<input checked="" type="checkbox"/>	query	((causes OR pneumonia OR infections) AND ("nov...

Body Cookies Headers (7) Test Results Status: 200 OK

Pretty Raw Preview Visualize JSON

```

34
35
36
37
38
39
40
41
    "KeywordInContext": [
      "antiviral agents",
      "virus infection",
      "selective inhibition",
      "side effect"
    ],
    "TextParagraph": "An effective selective antiviral agent with
selective inhibiting effect on a specific viral function o
therefore one object of the present invention to provide
  
```

GET <https://passageretrieval-api.artificialresearcher.com/search?domain=patent&simpleAuth=<api key>&userI> Send

Params ● Authorization Headers (6) Body Pre-request Script Tests Settings

<input checked="" type="checkbox"/>	domain	patent
<input checked="" type="checkbox"/>	simpleAuth	<api key>
<input checked="" type="checkbox"/>	userDefined	false
<input checked="" type="checkbox"/>	query	favipiravir spike protein RNA-dependent RNA pol...

Body Cookies Headers (7) Test Results Status: 200 OK Time: 8.90 s Size: 2.14 MB

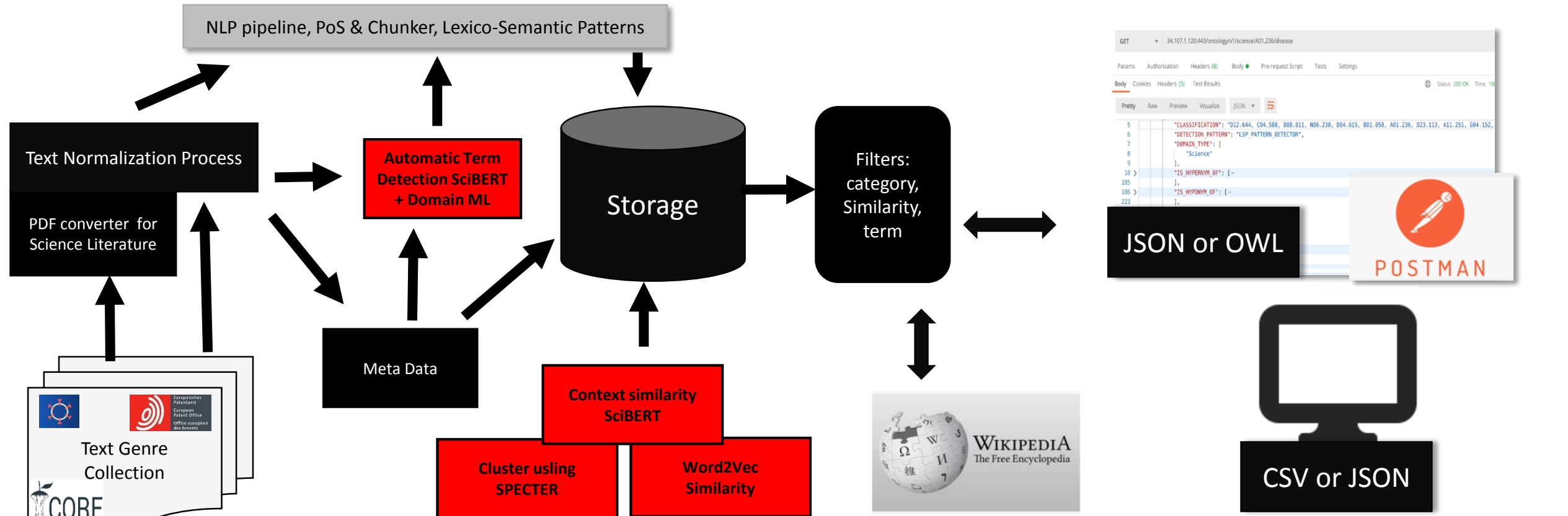
Pretty Raw Preview Visualize JSON

```

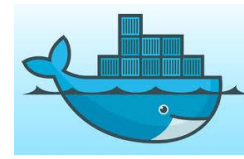
1
2
3
4
5
6
7
8
9
10
11
12
13
14
    {
      "Title": "\" - SUBSTITUTED CARBA-NUCLEOSIDE ANALOGS FOR ANTIVIRAL TREATMENT\"",
      "Link": "https://worldwide.espacenet.com/patent/search?q=pn%3DEP3210993A1",
      "Affiliation": [ ...
    ],
    "Authors": [ ...
    ],
    "PublicationDate": "2017-08-30T00:00:00Z",
  
```


Artificial Researcher Ontology Services

Artificial Researcher Ontology Service

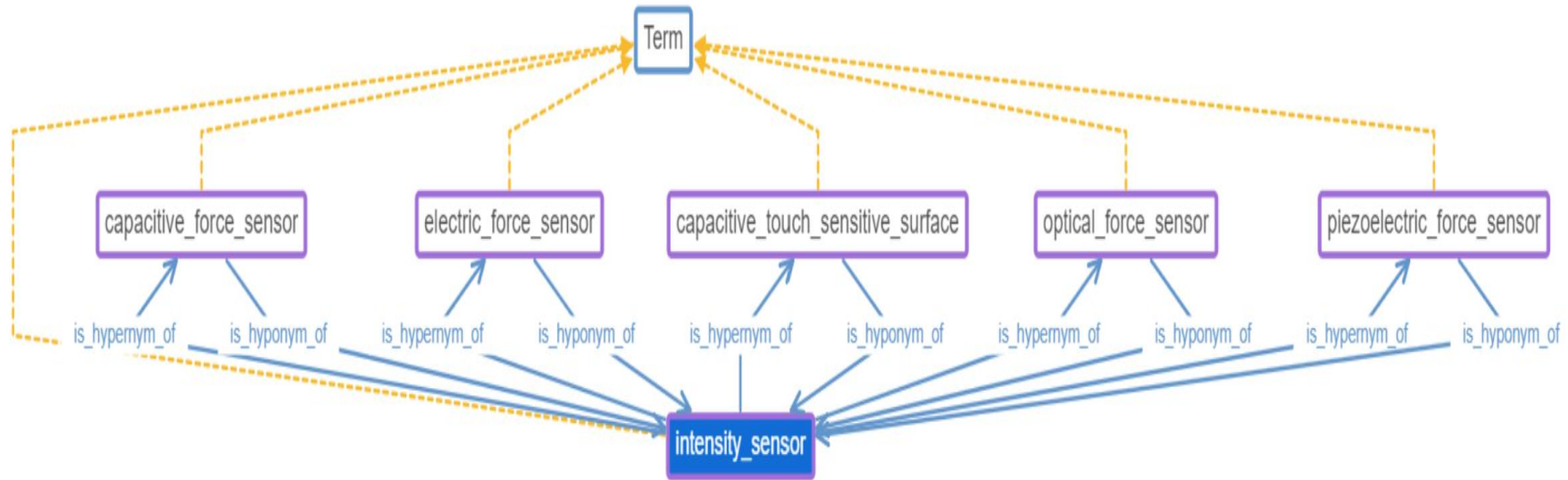


Open source Software Language Open source container



Relation Extraction Graph

Narrow terms




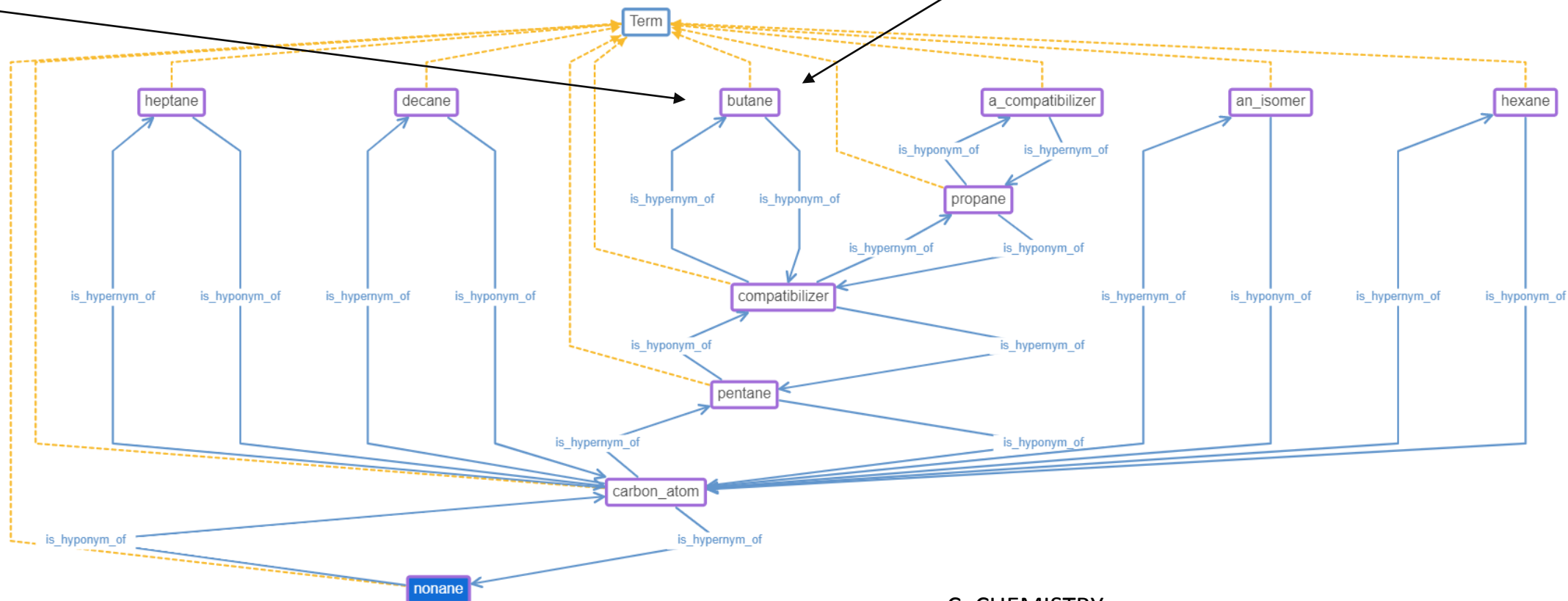
Broad term

Added Meta data to RE Graph

Label: organic compound

Wikipedia

Butane	
	
	
Names	
Preferred IUPAC name	Butane ^[1]
Systematic IUPAC name	Tetracarbane (never recommended ^[3])
Other names	Butyl hydride, ^[1] Quarlane, ^[2] Refrigerant 3-11-0
Identifiers	
CAS Number	106-97-8 ^g ✓
3D model (JSmol)	Interactive image ^g
Beilstein Reference	969129
ChEBI	ChEBI:37608 ^g ✓
ChEMBL	ChEMBL134702 ^g ✓
ChemSpider	7555 ^g ✓
ECOA InfoCard	100.003.136 ^g
EC Number	203-448-7
E number	E943a (glazing agents, ...)
Smolin Reference	1148
KEGG	D03186 ^g ✓
MeSH	butane ^g
PubChem CID	7843 ^g
RTECS number	EJ420000
UNII	6LV4FOR43R ^g ✓
UN number	1011
CompTox Dashboard (EPA)	DTXSID7024665 ^g ✓
InChI	[show]
SMILES	[show]
Properties	
Chemical formula	C ₄ H ₁₀
Molar mass	58.124 g·mol ⁻¹
Appearance	Colorless gas
Odor	Gasoline-like or natural gas-like ^[1]
Density	2.48 kgm ⁻³ (at 15 °C (59 °F))
Melting point	-140 to -134 °C; -220 to -209 °F; 133 to 139 K
Boiling point	-1 to 1 °C; 30 to 34 °F; 272 to 274 K
Solubility in water	61 mg L ⁻¹ (at 20 °C (68 °F))
log P	2.745
Vapor pressure	-170 kPa at 293 K ^[4]
Henry's law constant (K _H)	11 mmol Pa ⁻¹ kg ⁻¹
Conjugate acid	Butanium
Magnetic susceptibility (χ)	-57.4 · 10 ⁻⁶ cm ³ /mol
Thermochemistry	
Heat capacity (C)	98.49 J K ⁻¹ mol ⁻¹
Std enthalpy of formation (Δ _f H [∘] ₂₉₈)	-126.3—-124.9 kJ mol ⁻¹
Std enthalpy of combustion (Δ _c H [∘] ₂₉₈)	-2.8781—-2.8769 MJ mol ⁻¹
Hazards ^[6]	



- C: CHEMISTRY
 - C07: ORGANIC CHEMISTRY
 - C07F7/08; C07F7/10; C07F7/28
 - C08: ORGANIC MACROMOLECULAR COMPOUNDS
 - C08F4/58

Combining NLP & Distributional Semantics

Embedding identifies similarities between different words

- Underwear **similar to** underpants , undergarment, panties, underclothes
- Strength **similar to** strengths, strength, toughness, stronger, sfrength

Technical semantic relations are a mixture of single words and phrases

$$JoinedSimilarity = \sum_{\substack{i,j=1,n \\ i \neq j \\ i < j}}^N \frac{\cos\left(\vec{w}_i, \vec{w}_j\right)}{N}$$

- *synthetic fibers* **synonym to** *polyester fibers*
- *thrips* **hypernym to** *bulb fly larvae*

- w_i, w_j represent each word vector pair cosine similarity of a MWT
- N is the number of words for a MWT (Andersson et al 2017)

Joined Similarity

Syntagmatic

- Does “Network lan” and “communication link” have (hyponymy) relation? Yes



Paradigmatic

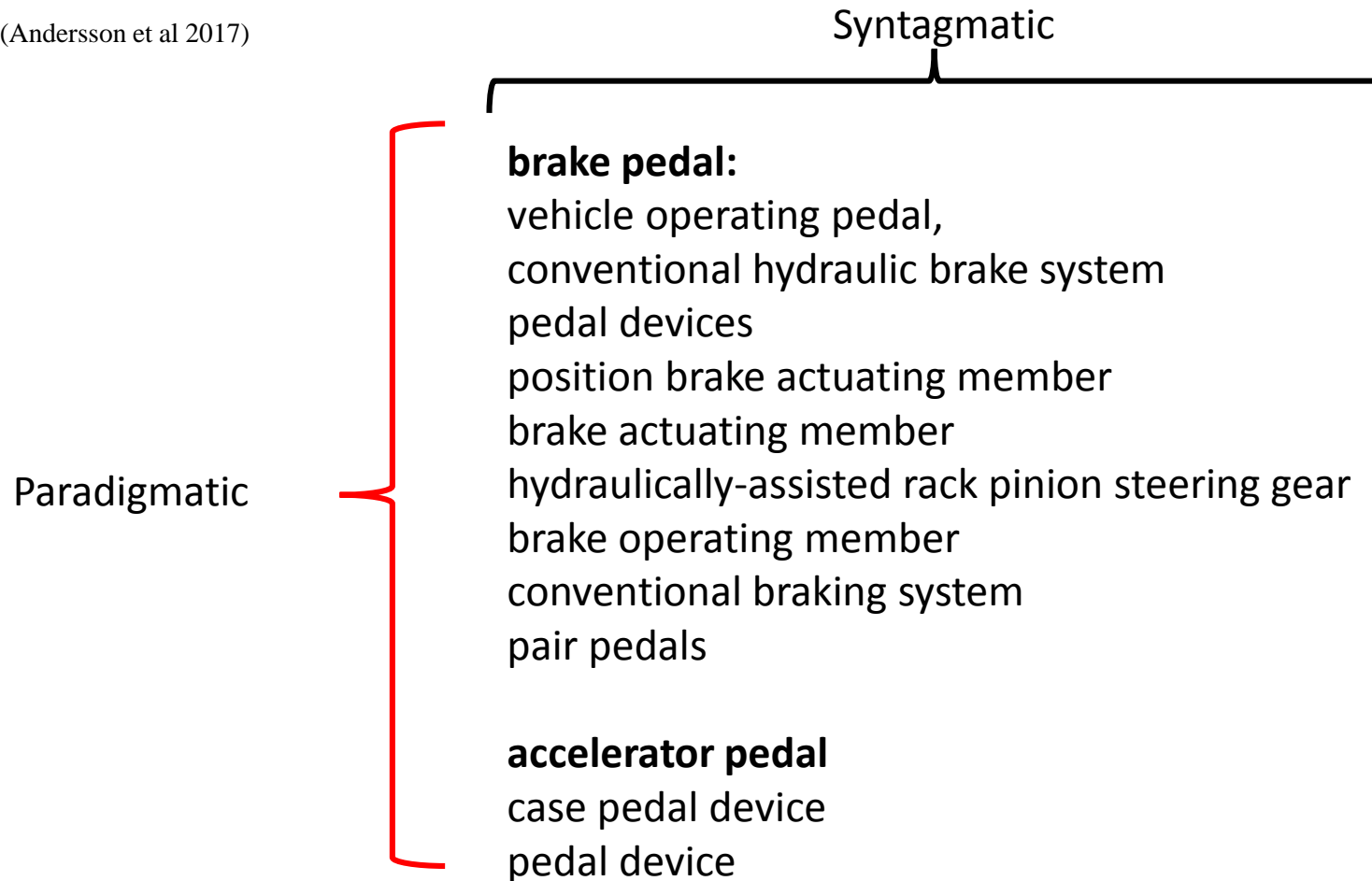
- Does “mechanical stress” and “communication link” have a (hyponymy) relation? No

$$JoinedSimilarity = \sum_{\substack{i,j=1,n \\ i \neq j \\ i < j}}^N \frac{\cos\left(\vec{w}_i, \vec{w}_j\right)}{N}$$

- w_i, w_j represent each word vector pair cosine similarity of a MWT
- N is the number of words for a MWT (Andersson et al 2017)

Identification of technical term and related concepts

(Andersson et al 2017)



Request via Postman

http://swagger.artificialresearcher.com

GET 34.107.1.120:443/ontology/v1/science/A01.236/disease

Params Authorization Headers (8) **Body** Pre-request Script Tests Settings

none form-data x-www-form-urlencoded raw binary GraphQL JSON

```

1 {
2   "api_key": "insert api key",
3   "return_type": "owl",
4   "window_size": 1,
5   "stride": 0

```

Body Cookies Headers (11) Test Results Status: 200 OK

Pretty Raw Preview Visualize Text

```

1 <?xml version="1.0"?>
2 <rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
3   xmlns:xsd="http://www.w3.org/2001/XMLSchema#"
4   xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
5   xmlns:owl="http://www.w3.org/2002/07/owl#"
6   xml:base="http://exp.artificialresearcher.com/ontology/demo/A01.236.owl"
7   xmlns="http://exp.artificialresearcher.com/ontology/demo/A01.236.owl#">

```

GET 34.107.1.120:443/ontology/v1/science/A01.236/disease

Params Authorization Headers (8) Body Pre-request Script Tests Settings

<input checked="" type="checkbox"/>	Postman-Token	<calculated when request is sent>
<input checked="" type="checkbox"/>	Content-Type	application/json
<input checked="" type="checkbox"/>	Content-Length	<calculated when request is sent>
<input checked="" type="checkbox"/>	Host	<calculated when request is sent>
<input checked="" type="checkbox"/>	User-Agent	PostmanRuntime/7.26.8
<input checked="" type="checkbox"/>	Accept	/*/*
<input checked="" type="checkbox"/>	Accept-Encoding	gzip, deflate, br
<input checked="" type="checkbox"/>	Connection	keep-alive

```

<owl:Class rdf:about="#Term">
  <rdfs:subClassOf rdf:resource="http://www.w3.org/2002/07/owl#Thing" />
</owl:Class>

<Term rdf:about="#disease">
  <rdf:type rdf:resource="http://www.w3.org/2002/07/owl#NamedIndividual" />
  <IS_HYPERNYM_OF rdf:resource="#polio" />
  <IS_HYPERNYM_OF rdf:resource="#vascular_" />
  <IS_HYPERNYM_OF rdf:resource="#malaria" />
  <IS_HYPERNYM_OF rdf:resource="#cholera" />
  <IS_HYPERNYM_OF rdf:resource="#lassa" />
  <IS_HYPERNYM_OF rdf:resource="#fever" />
  <IS_HYPERNYM_OF rdf:resource="#anosmia" />
  <IS_HYPERNYM_OF rdf:resource="#allergy" />
  <IS_HYPERNYM_OF rdf:resource="#aspergillosis" />
  <IS_HYPERNYM_OF rdf:resource="#tuberculosis" />
  <IS_HYPERNYM_OF rdf:resource="#leprosy" />
  <IS_HYPERNYM_OF rdf:resource="#influenza" />
  <IS_HYPERNYM_OF rdf:resource="#smallpox" />
  <IS_HYPERNYM_OF rdf:resource="#chikungunya" />
  <IS_HYPERNYM_OF rdf:resource="#measle" />

```


Artificial Research NLP Toolkit

Release First Quarter 2021

- SciPatentBERT Multi Word Term Detection
 - Alpha already released (Fink et al 2019)
 - Automatic Domain Term Recognition
 - MWT and single terms
- SciBERT text segment similarities
 - Alpha already released (Pretrained language model in Beltagy et 2019)
- COVID PIC identification
 - Population Intervention Comparison
- Patent SPECTER

Request via Postman

Text segment similarities

The screenshot displays the Postman interface for a POST request. The URL is `http://35.204.138.97:8080/predict?key=<key>`. The request body is a JSON object with a `texts` array containing a long string of text. The response status is `200 OK` with a time of `698 ms` and a size of `168 B`. The response body shows a similarity score of `0.8384154`.

Request Headers (11):


<input type="checkbox"/>	Content-Type	application/json
<input checked="" type="checkbox"/>	Content-Length	<calculated when request is sent>
<input checked="" type="checkbox"/>	Host	<calculated when request is sent>
<input checked="" type="checkbox"/>	User-Agent	PostmanRuntime/7.26.8
<input checked="" type="checkbox"/>	Accept	*/*
<input checked="" type="checkbox"/>	Accept-Encoding	gzip, deflate, br
<input checked="" type="checkbox"/>	Connection	keep-alive
<input checked="" type="checkbox"/>	key	insert the api key
<input checked="" type="checkbox"/>	Content-Type	application/json

Request Body:

```
1 [{"texts":["For other antiviral lectins such as concanavalin A , banana lectin and cyanovirin - N ( CV - N ) interactions between the lectin and as yet undescribed cellular moieties have been reported to induce un side effects including secretion of inflammatory cytokines and activation of host T - cells .","ACE2 is reduced in critical tissues such as lungs ( 24 ) , however , we can not suggest that low levels of ACE2 expression imply no effect of the virus on the tissue "]}]
```

Response Body:

```
1 0.8384154
```

A group of four women are smiling and looking towards the camera. They are positioned in front of a brick wall. The woman on the far left has dark hair and is wearing glasses and a dark top with a light-colored scarf. The woman next to her has dark hair and is wearing a dark top. The woman in the back center has blonde hair and is wearing a dark top. The woman on the far right has red hair and is wearing a patterned top. The text is overlaid in the center of the image.

For more information about the
company and the team
please visit
www.artificialresearcher.com

Reference

- Andersson L., Lupu M., Hanbury A. (2013) *Domain Adaptation of General Natural Language Processing Tools for a Patent Claim Visualization System*. In Proceedings of Multidisciplinary Information Retrieval, Eds. Mihai Lupu, Evangelos Kanoulas, Fernando Loizides, Lecture Notes in Computer Science, Springer Berlin Heidelberg, (70-82)
- Andersson, L., Lupu, M., Palotti, J., Hanbury, A., and Andreas, R. (2016) *When is the time Ripe for Natural Language Processing for Patent Passage Retrieval?*. In Proceedings of the 25th ACM International on Conference on Information and Knowledge Management, CIKM16.
- Andersson L., Hanbury A., Rauber A. (2017) *The Portability of three type of Text Mining Techniques into the patent text genre*. In M. Lupu, K. Mayer, J. Tait, and A. J. Trippe, Second edition, Current Challenges in Patent Information Retrieval
- Andersson L., Rekabsaz N., Hanbury A. (2017) *Automatic query expansion for patent passage retrieval using paradigmatic and syntagmatic information* The first WiNLP Workshop co-located with with the Annual Meeting of the Association for Computational Linguistics (ACL 2017), Vancouver
- Andersson L., (2021*) *The last two decades of Natural Language Processing in the the Intellectual Property Domain, VSI: Text Mining and Semantic Technologies in the Intellectual Property Domain*, World Patent Information (Guest Editorial)
- Beltagy, I., Lo K., and Cohan A. (2019) *SciBERT: A pretrained language model for scientific text*. arXiv preprint arXiv:1903.10676.
- Fink T., Andersson L., Hanbury A. (2019) *Detecting Multi Word Terms in Patents the same way as Named Entities*. In Proceeding 1st PatentSemTech Workshop, (Extended Abstract)
- Wang, L.L., Lo, K., Chandrasekhar, Y., Reas, R., Yang, J., Eide, D., Funk, K., Kinney, R.M., Liu, Z., Merrill, W., Mooney, P., Murdick, D., Rishi, D., Sheehan, J., Shen, Z., Stilson, B., Wade, A.D., Wang, K., Wilhelm, C., Xie, B., Raymond, D., Weld, D.S., Etzioni, O., & Kohlmeier, S. (2020). *CORD-19: The Covid-19 Open Research Dataset*. ArXiv.