



Kairntech / SCAI Webinar

“Automatic relation extraction from biomedical content for knowledge graph creation and maintenance”

June 8, 2022

info@kairntech.com

Welcome! Agenda:

- Introduction: Your hosts for today:



Prof. Martin Hofmann-Apitius
Head of Bioinformatics
Fraunhofer-SCAI
St. Augustin



Stefan Geißler
Co-founder, COO
Kairntech
Grenoble, Neckargemünd

- Construction of Knowledge Graphs on biomedical topics (Martin Hofmann-Apitius, 20min)
- AI powered content analysis (Stefan Geißler, 20min)
- Larger context, wrapping up and next steps (Martin Hofmann-Apitius, 10min)

- Q&A, Discussion, next steps, (all, 30min)

Introducing Kairntech

- 1 French NLP/AI startup created in December 2018
- 2 Grenoble, France (headquarter), Paris and Heidelberg, Germany
- 3 Selected customers: Boehringer-Ingelheim, SCAI-Fraunhofer, AFP, French Government, Groupe Revue Fiduciaire,...
- 4 Mission: Making NLP / Machine Learning accessible for domain experts (“No code / Low code platform”)

Project Intro & Background

- Spring 2021: SCAI: “We are looking for an automatic procedure to **extract** information about entities and their **relations** from a larger corpus on psychiatric disorders and render that in the BEL language. Can you do that?”
- Kairntech: “Sure!”
- Implementation phase until Sept 2021, assessment by SCAI
- Since then
 - Defined future cooperation in MOU
 - Joint publication efforts:
<https://www.biorxiv.org/content/10.1101/2022.03.07.483233v1>
 - Deployed Kairntech software at large SCAI High Performance Computing cluster
 - Preparing joint activities

Kairntech Studio (build a pipeline)

Import documents

TXT, PDF, Word, HTML, XML, JSON...

Create a training dataset

Explore, label text with manual or assisted **text annotation tools**

Create and compare learning models

Using built-in and state-of-the-art **algorithms**.

Create NLP pipelines

Combining **models**, client **taxonomy**, **Knowledge graphs**, technical **components**...

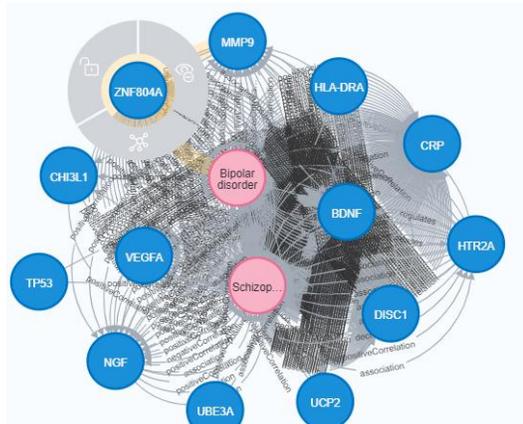
Usage via **No-code & easy-to-use web application**

or

via **RestAPI**

The screenshot displays the Kairntech Studio web application interface. On the left, there is a sidebar with a search bar and a list of filters for document selection, including 'Inside dataset', 'Label', 'Created by', 'Annotation date', 'Date of import', 'Imported by', 'lang', 'nature', 'version', and 'origin'. The main content area shows a list of documents with their titles and snippets of text. Each snippet has colored boxes highlighting specific entities, such as 'Spain', 'Calvino', 'IMF', 'GfK', 'Rolf Buerkl', 'Bibaa Henry', 'Nicole Smallman', 'Fryent Country Park', 'Wembley Park', 'London', 'Jean-Michel Aulas', 'Nations', 'Saskatchewan', 'CBC', 'Cohen', 'Farah Obaidullah', 'Deep Sea Conservation Coalition', and 'Baerbock'. Below the document list, there is a 'Machine learning experiments' table with columns for 'Name', 'Status', 'F1 score', 'Precision', 'Recall', 'Created', and 'Updated'. A 'Quality report' is also visible, showing a bar chart of 'Average and per-label F-measure' for various labels like 'Loc-Org' and 'Location'. The bottom of the interface shows a footer with 'Made with © by Kairntech' and '© Kairntech 2022'.

Kairntech Server (run the pipeline in production)



Knowledge applications



REST API

NLP Pipeline for SCAI

Link entities to
WikiData taxonomies

- Large-scale entity recognition
- Multi-topic
- Multi-lingual
- Constantly updated
- Linked to background information

Refine entity
extraction

Custom housekeeping
to improve
recognition quality:
Filtering

Compute add'tl
entity properties

Proteins receive label
about protein
modification (ex:
phosphorylation)

Build relations
between entities

Apply custom-
trained relation
extraction model

Format results
to output

Return results in
BEL format for
integration in SCAI
Knowledge Graph
environment

Entity Linking & Namespaces

Kairntech entity extraction off-the-shelf ready to cover this and many other requirements. Here nothing had to be built for the project.

A subset of namespaces is used:

- [HGNC](#) for proteins
- [MeSH](#) for pathologies
- [CheBI](#) for drug/chemical abundances
- [GO](#) for biological processes

Recognized entities are

- Disambiguated (“cancer”: animal or disease?)
- Scored (if the concept at the core of a document’s topic or peripheral?)
- Normalized (synonyms → preferred form, ex: “NIDDM” → “Diabetes Mellitus Type 2”)
- Linked: Entities are linked to world-knowledge

NLP Pipeline for Entity Recognition & Linking: Example!

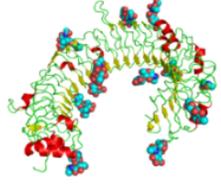
Kairntech Sherpa SCAI PMOD

Test

Syncytin-1, an endogenous retroviral protein, triggers the activation of CRP via TLR3 signal cascade in glial cells. Schizophrenia is a devastating psychiatric disorder that impacts on social functioning and quality of life, and there is accumulating evidence that inflammation is a potential pathogenic mechanism of schizophrenia. However, the mechanism of inflammation possibly occurred in schizophrenia has not been well understood. The endogenous retroviral protein syncytin-1 and inflammatory marker CRP are both abnormally expressed in schizophrenia patients. CRP is one of the markers of bacterial infection generally. Less clear is whether virus or viral protein can trigger the activation of CRP. Here, we detected a robust increase of the levels of syncytin-1 and CRP in schizophrenia patients, and displayed a positive correlation and marked consistency between expressions of syncytin-1 and CRP in schizophrenia patients. Furthermore, overexpression of syncytin-1 significantly elevated the levels of CRP, TLR3, and IL-6 in both human microglia and astrocytes. TLR3 deficiency impaired the expressions of CRP and IL-6 induced by syncytin-1. Importantly, we observed a cellular co-localization and a direct interaction between syncytin-1 and TLR3. Additionally, knockdown of IL-6 inhibited the syncytin-1 induced CRP which might explain the elevated schizophrenia

```
144      "wikipediaExternalRef": "307809"
145    }
146  },
147  ],
148  },
149  {
150    "start": 81,
151    "end": 85,
152    "labelName": "protein",
153    "label": "Protein",
154    "text": "TLR3",
155    "score": 0.791,
156    "properties": {
157      "namespace": "HGNC",
158      "name": "TLR3",
159      "identifier": "11849"
160    },
161    "terms": [
162      {
163        "identifier": "Q14866230",
164        "lexicon": "wikidata",
165        "preferredForm": "TLR3",
166        "score": 1,
167        "properties": {
168          "wikidataId": "Q14866230",
169          "wikipediaExternalRef": "2082102"
170        }
171      }
172    ]
173  }
174  ]
175  }
```

TLR3



Toll-like receptor 3 (TLR3) also known as CD283 ([cluster of differentiation 283](#)) is a [protein](#) that in humans is encoded by the "TLR3" [gene](#). TLR3 is a member of the [toll-like receptor](#) family of [pattern recognition receptor](#) of the [innate immune system](#). TLR3 recognizes double-stranded [RNA](#) in [endosomes](#), which is a common feature of [viral genomes](#) internalised by [macrophages](#) and [dendritic cell](#).

Wikidata statements

Wikipedia Wikidata

Made with  by Kairntech

Kairntech 2022

Relationship extraction

- Trained on a **training data set** provided by Fraunhofer SCAI
- The table (right) lists the most important relations addressed in the project
- The model determines whether between any pair of entities a relation holds or not (“NoRelation”) and if yes, which one
- Quality decreases with smaller numbers of training examples per relationship type (as expected).
- Manual inspection of (a sample of) the results from Kairntech by SCAI experts: ~73% of the relations returned by Kairntech are valid

RELATION NoRelation

RELATION increases

RELATION decreases

RELATION regulates

RELATION positiveCorrelation

RELATION association

RELATION negativeCorrelation

Complete NLP Pipeline with Relationship Extraction: Example!



KAIRNTECH Sherpa

Demonstration page

Sherpa URL
https://sherpa-scw.kairntech.com/

Name
demo

Password

Connect

Select project
Scai Pmod

Select annotator
BEL Relations

Built with Streamlit and sherpa-scw

Input text to analyze

Syncytin-1, an endogenous retroviral protein, triggers the activation of CRP via TLR3 signal cascade in glial cells. Schizophrenia is a devastating psychiatric disorder that impacts on social functioning and quality of life, and there is accumulating evidence that inflammation is a potential pathogenic mechanism of schizophrenia. However, the mechanism of inflammation possibly occurred in schizophrenia has not been well understood. The endogenous retroviral protein syncytin-1 and inflammatory marker CRP are both abnormally expressed in schizophrenia patients. CRP is one of the markers of bacterial infection generally. Less clear is whether virus or viral protein can trigger the activation of CRP. Here, we detected a robust increase of the levels of syncytin-1 and CRP in schizophrenia patients, and displayed a positive correlation and marked consistency between expressions of syncytin-1 and CRP in schizophrenia patients. Furthermore, overexpression of syncytin-1 significantly elevated the levels of CRP, TLR3, and IL-6 in both human microglia and astrocytes. TLR3 deficiency impaired the expressions of CRP and IL-6 induced by syncytin-1. Importantly, we observed a cellular co-localization and a direct interaction between syncytin-1 and TLR3. Additionally, knockdown of IL-6 inhibited the syncytin-1-induced CRP expression. Thus, the totality of these results showed that viral protein syncytin-1 could trigger the activation of CRP, which might explain the elevated CRP in sterile inflammation and exhibit a novel mechanism for regulation of inflammation by syncytin-1 in schizophrenia.

Process Text

Or upload text/json file to analyze

Drag and drop file here
Limit 200MB per file

Browse files

Process File

Annotation & formatting successful!

Download result

Table

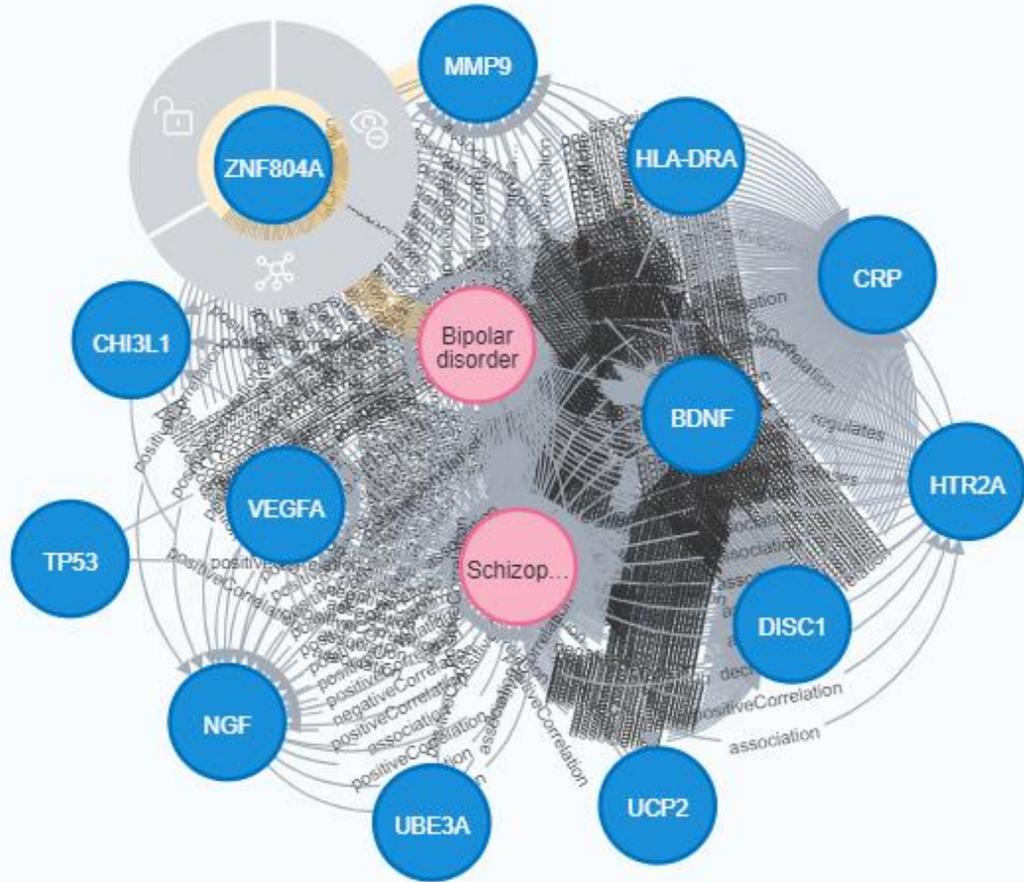
Sheet1

	Subject	Subject.start	Subject.end	Subject.text	Relation	Object	Object.start	Object.end	Object.text	Score	PubMedID	EvidenceSentence	PublicationTitle	Journal
0	p(HGNC:13525 ! ERVW-1)	0	10	Syncytin-1	increases	p(HGNC:2367 ! CRP)	73	76	CRP	0.9514	<NA>	Syncytin-1, an endogenous...	<NA>	<NA>
1	p(HGNC:11849 ! TLR3)	81	85	TLR3	increases	p(HGNC:2367 ! CRP)	73	76	CRP	0.8791	<NA>	Syncytin-1, an endogenous...	<NA>	<NA>
2	path(MESH:D012559 ! sch...)	106	119	schizophrenia	positiveCorrelation	p(HGNC:13525 ! ERVW-1)	34	44	syncytin-1	0.8251	<NA>	The endogenous retroviral ...	<NA>	<NA>
3	path(MESH:D012559 ! sch...)	106	119	schizophrenia	positiveCorrelation	p(HGNC:2367 ! CRP)	69	72	CRP	0.8706	<NA>	The endogenous retroviral ...	<NA>	<NA>
4	path(MESH:D012559 ! sch...)	75	88	schizophrenia	positiveCorrelation	p(HGNC:13525 ! ERVW-1)	53	63	syncytin-1	0.9507	<NA>	Here, we detected a robust ...	<NA>	<NA>
5	p(HGNC:6018 ! IL6)	27	31	IL-6	increases	p(HGNC:2367 ! CRP)	65	68	CRP	0.9419	<NA>	Additionally, knockdown o...	<NA>	<NA>
6	p(HGNC:13525 ! ERVW-1)	46	56	syncytin-1	increases	p(HGNC:2367 ! CRP)	65	68	CRP	0.9454	<NA>	Additionally, knockdown o...	<NA>	<NA>
7	p(HGNC:13525 ! ERVW-1)	62	72	syncytin-1	increases	p(HGNC:2367 ! CRP)	105	108	CRP	0.9561	<NA>	Thus, the totality of these r...	<NA>	<NA>
8	p(HGNC:13525 ! ERVW-1)	62	72	syncytin-1	increases	p(HGNC:2367 ! CRP)	143	146	CRP	0.9534	<NA>	Thus, the totality of these r...	<NA>	<NA>

Thus, the totality of these results showed that viral protein syncytin-1 could trigger the activation of CRP, which might explain the elevated CRP in sterile inflammation and exhibit a novel mechanism for regulation of inflammation by syncytin-1 in schizophrenia.

Results displayed as a Knowledge Graph: Example!

“Which proteins are known to be positively correlated both with Schizophrenia as well as Bipolar Disorders?”



Relationship Properties

positiveCorrelation

NIH National Library of Medicine
National Center for Biotechnology Information

PubMed.gov

Advanced

> JAMA Psychiatry. 2014 Oct;71(10):1112-20. doi: 10.1001/jamapsychiatry.2014.1079.

Expression of ZNF804A in human brain and alterations in schizophrenia, bipolar disorder, and major depressive disorder: a novel transcript fetally regulated by the psychosis risk variant rs1344706

Ran Tao ¹, Helena Cousijn ², Andrew E Jaffe ¹, Philip W J Burnet ², Freya Edwards ², Sharon L Eastwood ², Joo Heon Shin ¹, Tracy A Lane ², Mary A Walker ², Brady J Maher ¹, Daniel R Weinberger ¹, Paul J Harrison ², Thomas M Hyde ¹, Joel E Kleinman ¹

Affiliations + expand
PMID: 25162540 PMID: PMC5894803 DOI: 10.1001/jamapsychiatry.2014.1079

Free PMC article

Save Email

- Speed up the analysis by parallelizing the process: Make use of SCAI's high performance computing cluster, leverage SCAI expertise in parallelizing complex software processes
- Investigate extension to other therapeutic areas
- Define joint approach & offering for industry use cases on
 - computing topic specific knowledge graphs
 - Updating/extending knowledge graphs
- Expand approach to cover large chunks (all?) of Medline?

- Kairntech off-the-shelf entity extraction performs well even in this highly specific subdomain
- Kairntech notion of processing pipelines allows to define sophisticated processing chains (here: entity recognition, application of specific custom model (ModType), relation extraction, output encoding into BEL syntax)
- Relation results assessed and declared useful by SCAI experts after detailed manual analysis
- Relation extraction from large literature corpus to feed Knowledge Graphs is feasible