



Kairntech Use Case: A Contract Analysis NLP Pipeline - Screenshots

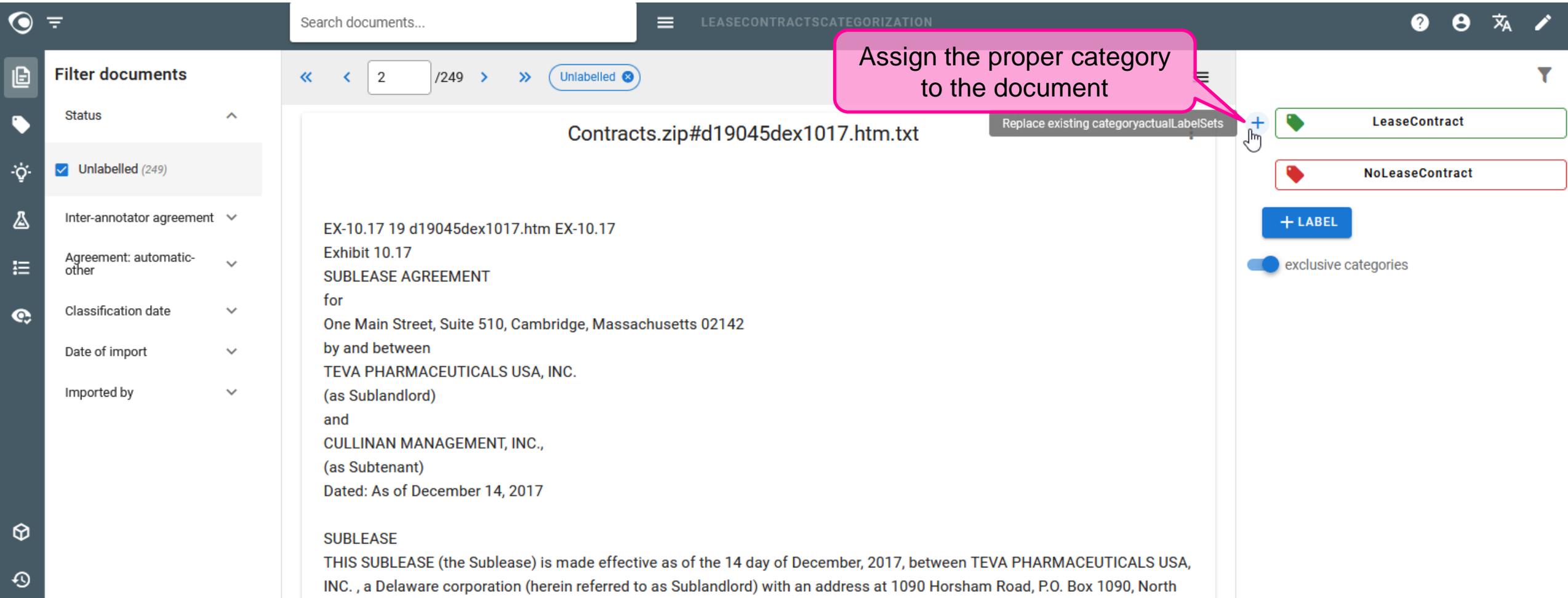
Kairntech webinar, Feb 28, 2023

info@kairntech.com

A Contract Analysis Scenario:

- ❖ We need to review a large number of Leasing Contracts: Who is leasing, since when and where?
- ❖ But the Leasing Contracts are part of a larger corpus of other documents
- ❖ So we need to first find the documents we want
- ❖ ... and then from these extract the required information
- ❖ Let's see how we do all this inside the Kairntech software, without writing a single line of code.

Manual labelling of documents



Search documents...

LEASECONTRACTSCATEGORIZATION

2 /249 Unlabelled

Contracts.zip#d19045dex1017.htm.txt

Replace existing categoryactualLabelSets

LeaseContract

NoLeaseContract

+ LABEL

exclusive categories

Assign the proper category to the document

Filter documents

Status

Unlabelled (249)

Inter-annotator agreement

Agreement: automatic-other

Classification date

Date of import

Imported by

EX-10.17 19 d19045dex1017.htm EX-10.17

Exhibit 10.17

SUBLEASE AGREEMENT

for

One Main Street, Suite 510, Cambridge, Massachusetts 02142

by and between

TEVA PHARMACEUTICALS USA, INC.

(as Sublandlord)

and

CULLINAN MANAGEMENT, INC.,

(as Subtenant)

Dated: As of December 14, 2017

SUBLEASE

THIS SUBLEASE (the Sublease) is made effective as of the 14 day of December, 2017, between TEVA PHARMACEUTICALS USA, INC. , a Delaware corporation (herein referred to as Sublandlord) with an address at 1090 Horsham Road, P.O. Box 1090, North

Train a model and apply it on the corpus

Model experiments

+ EXPERIMENT

Actions	Name	Engine	Options	F-measure
☆ ✕ ⋮ LAUNCH	LeaseNoLease	Sklearn	mongo_fromcol: documents, classifier: KerasMLPClassifier, shuffle: true	90.9 %
☆ ✕ ⋮ LAUNCH	Spacy	Spacy	mongo_fromcol: documents	0.0 %
☆ ✕ ⋮ LAUNCH	SVM	Sklearn	mongo_fromcol: documents, featurizer: tfidf, ngram_range: [1,3], classifier: LinearSVC	92.7 %
☆ ✕ ⋮ LAUNCH	SVM with scores	Sklearn	mongo_fromcol: documents, featurizer: tfidf, ngram_range: [1,3], classifier: SVC, kernel: linear	0.0 %
☆ ✕ ⋮ LAUNCH	XGBoost	Sklearn	mongo_fromcol: documents, featurizer: tfidf, ngram_range: [1,3], classifier: XGBClassifier	94.3 %

1 Once enough documents have been labelled (here 30-50 are enough) you can train an ML model

2 System provides you with quality feedback

Satisfied with the training success? Apply your new model on the rest of the corpus to automatically label all documents

Now add the required entities

The screenshot displays the KAIRNTECH interface for document extraction. The main window shows a document titled "Contracts.zip#a2242765zex-10_1.htm.txt" with the following text: "EX-10.1 11 a2242765zex-10_1.htm EX-10.1 Exhibit 10.1 AGREEMENT OF LEASE BY AND BETWEEN NOVAPARK LLC Landlord, and ANGION BIOMEDICA CORP. Tenant. Dated: June 21, 2011 Demised Premises: A portion of 51 Charles Lindberg Boulevard Uniondale, New York". Annotations include yellow boxes around "NOVAPARK LLC" and "ANGION BIOMEDICA CORP.", a green box around "June 21, 2011", and a red box around "51 Charles Lindberg Boulevard Uniondale, New York". A sidebar on the left titled "Filter documents" shows checkboxes for "Labelled (59)", "Unlabelled (24)", "ContractParty (52)", "Date (52)", and "ObjectAddress (32)". A right sidebar shows a list of labels: "ContractParty", "Date", and "ObjectAddress", with a "+ LABEL" button below. The top navigation bar includes a search bar, a menu icon, and the text "LEASECONTRACTSEXTRACTION".

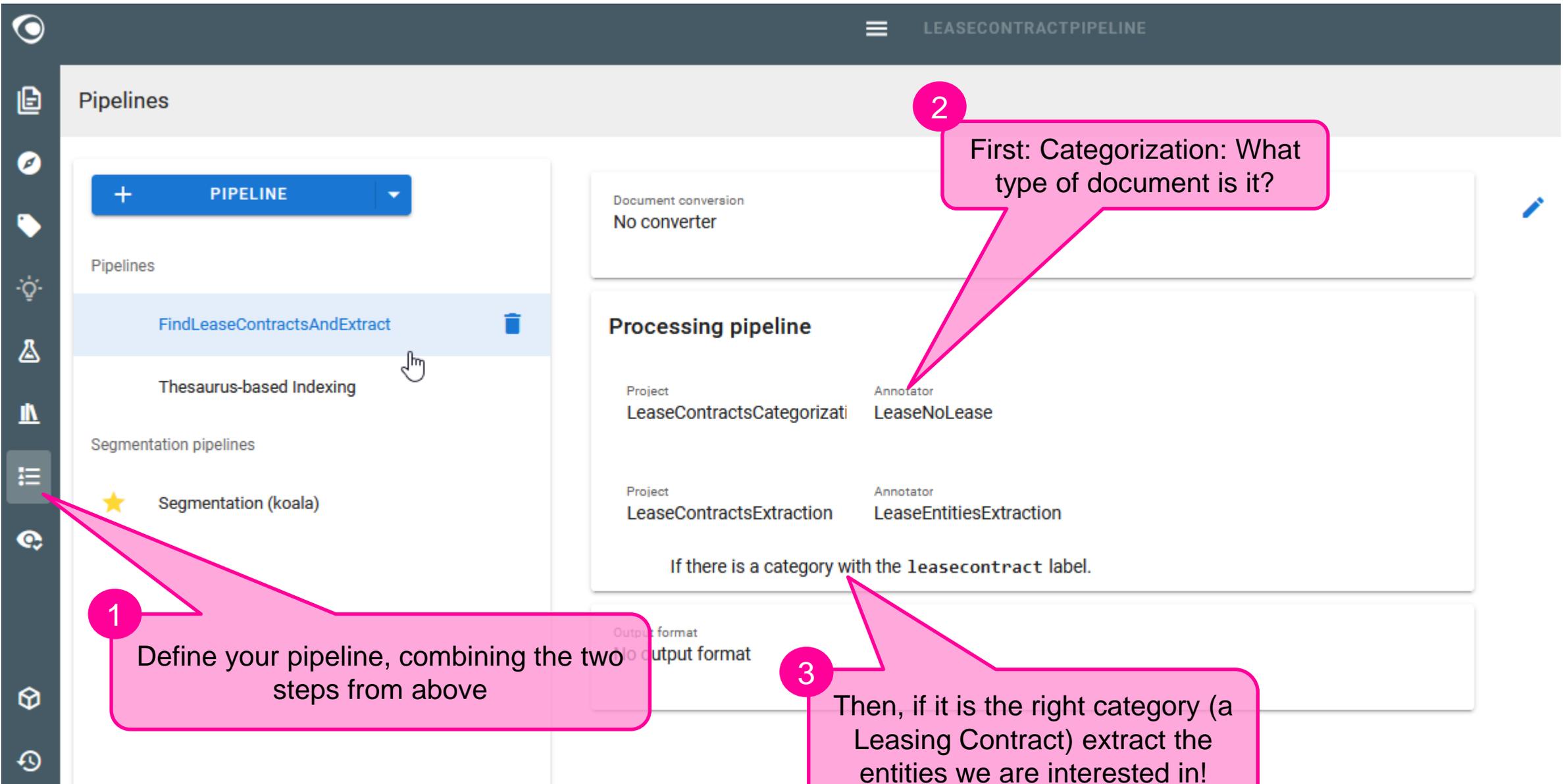
1 Do the same now on the selected documents: Define your required entities ...

2 ... and manually label the documents

3 The system learns in the background and after a while starts to generate new ones based on the provided samples, so the user can proceed quicker.

As before: On the labelled documents in the corpus, train an ML model, this time for Entity Extraction

Putting it all together



1 Define your pipeline, combining the two steps from above

2 First: Categorization: What type of document is it?

3 Then, if it is the right category (a Leasing Contract) extract the entities we are interested in!

Processing pipeline

Project	Annotator
LeaseContractsCategorizati	LeaseNoLease
LeaseContractsExtraction	LeaseEntitiesExtraction

If there is a category with the leasecontract label.

Testing the new pipeline.

1

This is certainly no Leasing Contract, so nothing to extract.

#noleasecontract (100%)

S-1/A 1 a2242585zs-1a.htm S-1/A

Use these links to rapidly review the document

TABLE OF CONTENTS

INDEX TO FINANCIAL STATEMENTS

As filed with the Securities and Exchange Commission on November 20, 2020.

Registration No. 333-249554

UNITED STATES

SECURITIES AND EXCHANGE COMMISSION

Washington, D.C. 20549

-

-

Amendment No. 2

to

FORM S-1

2

This however is a Leasing Contract, so the system extracts the desired entities

#leasecontract (100%)

EX-10.1 9 a2242620zex-10_1.htm EX-10.1

Exhibit 10.1

100 Binney, Cambridge, MA/Sigilon Therapeutics, Inc.

LEASE AGREEMENT

THIS LEASE AGREEMENT is made as of this 28 day of August, 2017, between ARE-MA REGION NO. 45, LLC, a Delaware limited liability company ('Landlord'), and SIGILON THERAPEUTICS, INC., a Delaware corporation ('Tenant').

BASIC LEASE PROVISIONS

Address:

100 Binney Street, Cambridge, Massachusetts.

Premises:

That portion of the Project, containing approximately 22,746 rentable square feet, located on the 6th floor and in designated portions of mechanical/equipment space located on the 1st floor and Level B-1 of the Building (as defined below), as shown on Exhibit A.



- ❖ Learn more about Kairntech and what we do
 - ❖ Website with use cases, blog, product description: www.kairntech.com
 - ❖ Product Video : <https://kairntech.com/doc/video/>
 - ❖ Documentation, FAQ: www.kairntech.com/doc

- ❖ Interested to try it out yourself? (no costs, no commitment)
 - ❖ Ask for a free trial www.kairntech.com

- ❖ Any suggestion, question, comment?
 - ❖ info@kairntech.com



*Thank you for your
attention!*

info@kairntech.com