

# Handbook: DeepAnnotation

## What is the DeepAnnotation Tool?

Many of the machine learning algorithms such as Natural Language Processing models or Question Answering specifically are in need of large amounts of labelled/ annotated data, in order for the model to function correctly. Labeled data are “pieces of data that have been tagged with one or more labels identifying certain properties or characteristics, or classifications or contained objects.”<sup>1</sup> In image recognition, labelled data could be a set of pictures where objects in the picture have already been marked and named or colours have been identified. In natural language processing, labels consist of document classifications (assigning a document the tag *report*, *contract*, *etc.*), entity recognition (assigning words within documents the tag *names*, *organizations*, *technologies*) or question-answer pairs for existing text. In the context of NLP Question Answering, large sets of question-answer pairs are needed as labels. Those pairs include a question, and then a specific text passage/ series of words (and its location) in one of the relevant documents that portrays the answer to the prior question.

**Generating enough of those question-answer pairs can be a lengthy and messy process. Therefore, the DeepAnnotation Tool offers a simple and intuitive interface for creating labels, for dividing workloads and for managing documents and labelling workforce. On the following pages, you will find all necessary information and tips for how to use the tool.**

---

<sup>1</sup> <https://www.techopedia.com/definition/33695/labeled-data>

# Table of Content

<b>General Remarks for Labelling</b>	<b>3</b>
<b>Your Role</b>	<b>3</b>
<b>PART 1 - THE ADMIN'S TASKS</b>	<b>3</b>
Setting up the tool	3
Creating a Project	4
Adding Documents	5
Preparing the Labelling Process	5
Creating Standard Questions	7
User Management	8
Exporting labels	9
<b>PART 2 - THE LABELLER'S TASK</b>	<b>9</b>
Navigating to the right questions	9
How to label	10
Labelling for Category "U"	11
<b>General remarks on labelling and question creation</b>	<b>12</b>

# General Remarks for Labelling

The quality of the algorithm using the labels highly depends on the quality and quantity of labels provided. Generally, the more labels the better as the model becomes more robust. Simultaneously, the more diverse the labels the better as the model becomes more flexible. In terms of the number of labels, you should aim at >2.000 question-answer pairs.

## Your Role

Your personal first steps towards using the tool depend on whether you are the admin/ project manager for the labelling process of your organization or of you are a labeller. If your task is to label data, you can skip the chapters on “setting up the tool” and jump directly to the instructions on “how to label”. If you are in an admin function, the next chapters will provide you with the information on how to set up the tool, including projects, documents and users.



**Attention:** As both Labeller or Admin, please use the Google Chrome Browser for using this tool!

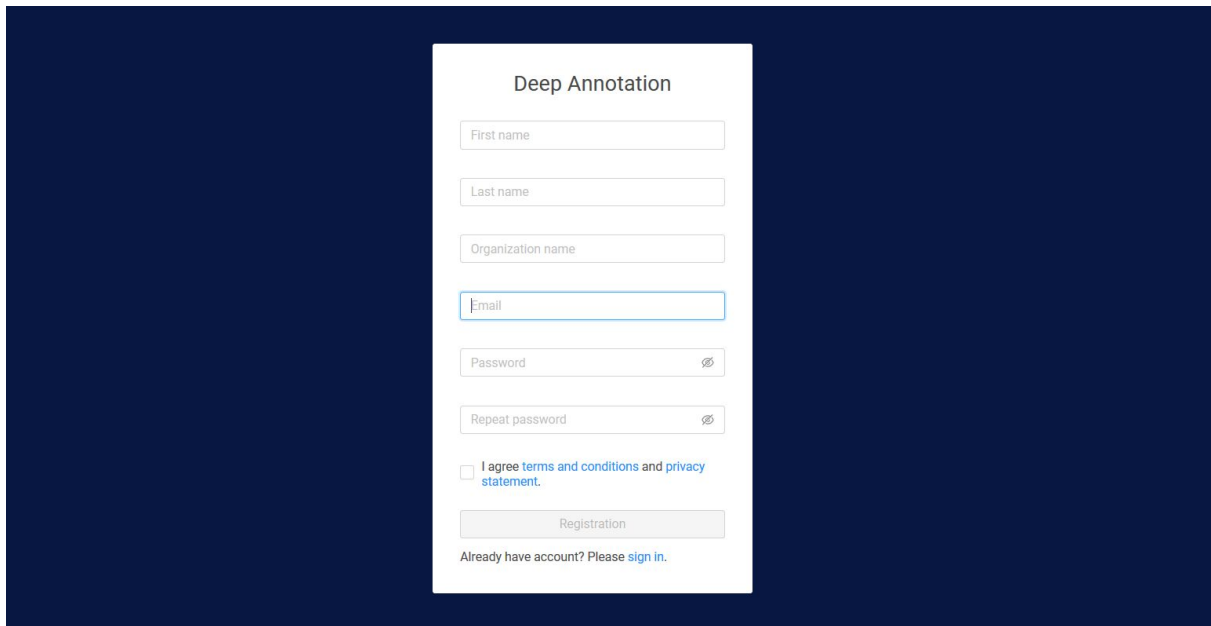
## PART 1 - THE ADMIN'S TASKS

### Setting up the tool

As a very first step, click the following link: “<https://annotate.deepset.ai/login>”. If you already have an account, simply log in with your email address and your password. Otherwise, click on “Please sign up here” and you will be forwarded to interface shown in Picture 1. Enter your name, organization, email address and set a password. After agreeing to the terms and conditions and privacy statement, all you need to do is click “Registration” and you are ready to go.



**Attention:** Each organization can only have one admin account. If you get the error that your organization already exists, someone else has already created the admin account. Please clarify the admin responsibility within your organization and contact us, if you need additional support.



Deep Annotation

First name

Last name

Organization name

Email

Password

Repeat password

☐ I agree [terms and conditions](#) and [privacy statement](#).

Registration

Already have account? Please [sign in](#).

Picture 1: Registration Mask

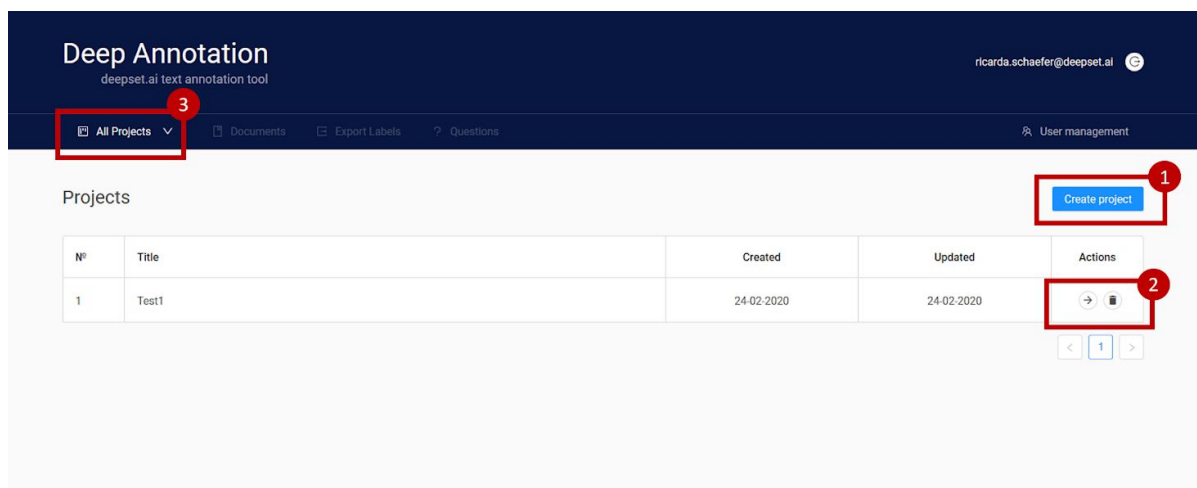
## Creating a Project

The first thing you will need to do, once you are logged in, is create a project. Click on the blue box in the upper right corner (Mark 1 in Picture 2) and simply enter your project name. After you have created the project, all your active projects will automatically appear on the main page. If you need to delete a project, click the small trash symbol in the “Action” column next to the project you want to delete (Mark 2 in Picture 2). In order to continue with the labelling of one of your projects, you need to select the respective project.



**Attention:** There are two ways to select a project for labelling:

- 1: Hover over “All Projects” in the upper left corner (Mark 3 in Picture 2) and click on the name of the project you want to proceed with.
- 2: Click on the little arrow in the “Actions” column (Mark 2 in Picture 2) right next to the project you want to proceed with.



Deep Annotation  
deepset.ai text annotation tool

ricarda.schaefer@deepset.ai

All Projects Documents Export Labels Questions User management

Projects

Create project

Nº	Title	Created	Updated	Actions
1	Test1	24-02-2020	24-02-2020	→ 🗑️

< 1 >

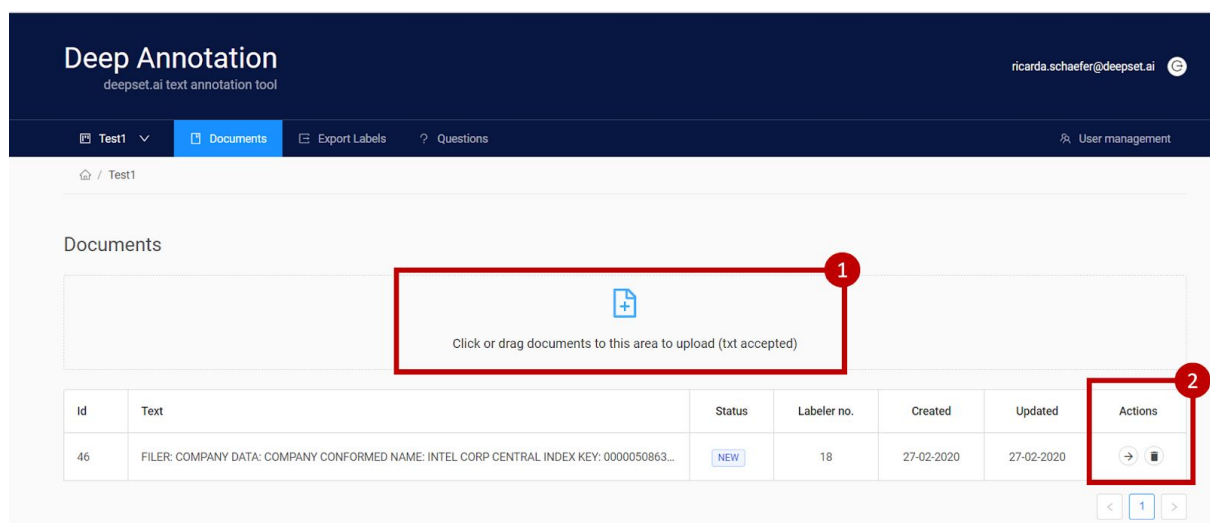
Picture 2: Create and manage projects

## Adding Documents

In a next step, you will need to upload the relevant documents that will be used for labelling. The accepted file format is .txt and all data in the tool needs to be cleaned from tables, images, headers etc. so that it is ready to be used for training right afterwards.

In order to add a document, drag it into the blue box or click in the box and then select the file from your computer (Mark 1 Picture 3). Afterwards, all uploaded documents will appear in the table below. In the “Actions” column (Mark 2 Picture 3) you can delete documents from the tool (trash symbol) or start the labelling process by clicking on the little arrow.

Please note that you do not need to upload all documents used for the actual QA later.



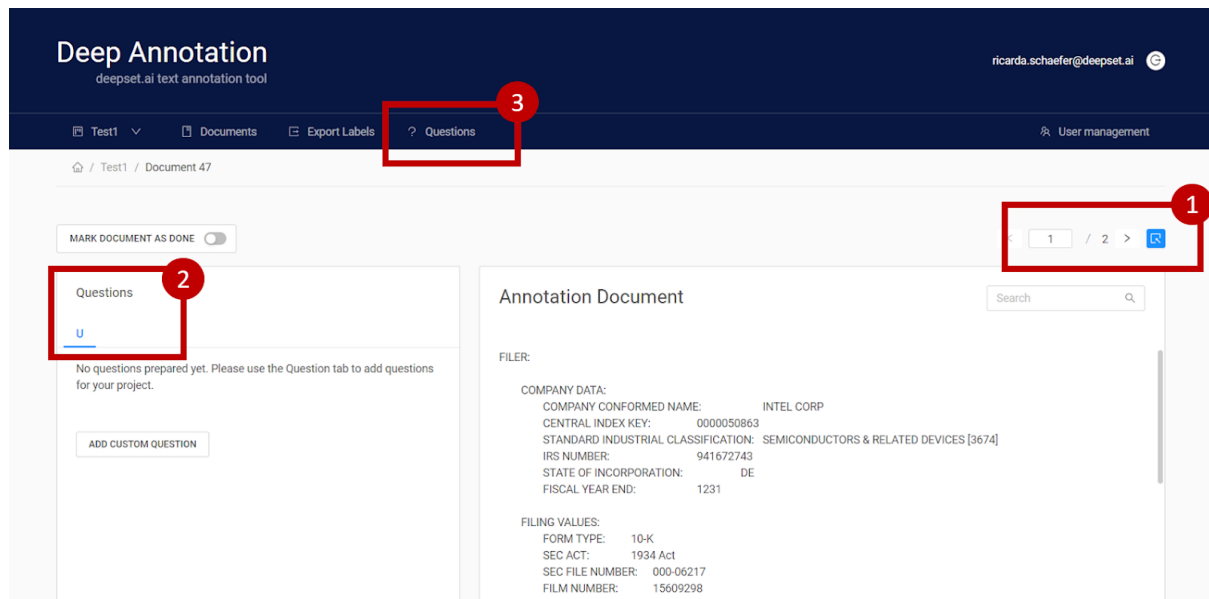
Picture 3: Uploading and managing documents



**Attention:** Technically, you can use any document (e.g. Wikipedia article) for the labelling process. However, if your QA model is very domain-specific (e.g. finance, law, engineering) it makes sense to use documents of that domain for labelling in order to train the model on the specific language.

## Preparing the Labelling Process

After you uploaded all the documents you want to prepare the labelling for and you have selected one document to start with, you will be forwarded to the following page:



**Picture 4: Document Labelling Page**

You can see the currently selected document on the right side and all available questions on the left side. To switch documents, either click the left and right arrow keys on your keyboard, or enter a document number/ click on the arrow on the upper right corner of the page (Mark 1 Picture 4).

Generally, the tool is designed to create blocks of work that can easily be assigned to different labellers. This is done to have a structured overview on who is working on what and to avoid overlaps. These blocks - called “categories” in this tool are named with capital letters. Right now, you can only see the category “U” in the “Question” field of the page (Mark 2 Picture 4). After you have created more blocks, they will appear next to the “U” on this page.

To understand your next tasks, you need to know that there are two types of questions in labelling question-answer pairs for QA:

1. **standard questions:** questions that will be labelled many times across different documents
2. **unique questions:** questions the labeler creates him-/herself and will most likely appear only once in the process

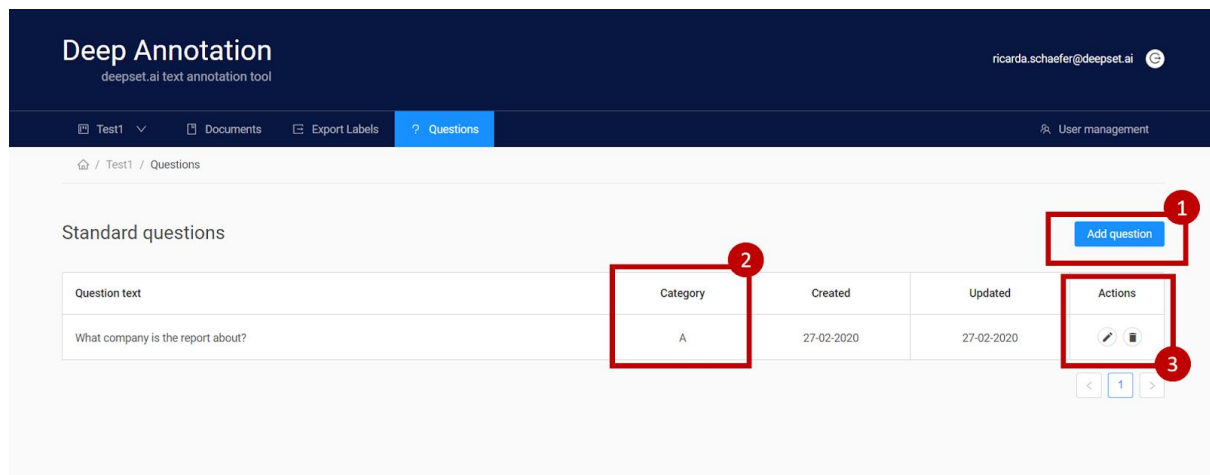
The block “U” you just saw is including all unique questions a labeler creates. All standard questions will be assigned to different blocks. As a project manager, you should prepare/ or delegate the preparation of blocks of standard questions before the process. These questions will then be given to the labelers who will search for the answers in the document. To do so, please click “Questions” (Mark 3 Picture 4) in the upper area of the page.

## Creating Standard Questions

Once you were directed the Questions page, you have the possibility to create, change and delete questions. First, create questions by clicking on the “Add question” button (Mark 1 Picture 5). Then enter the question text and assign the question to a working block/ category (Picture 6).

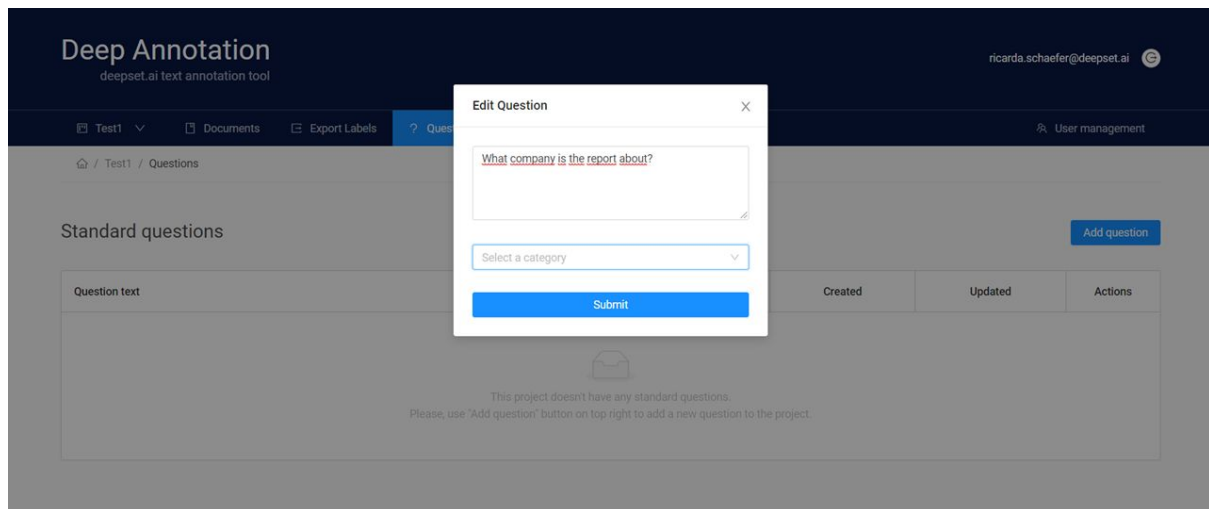
When creating the standard questions, try to consider a few things:

- Create questions that can be answered by the documents you uploaded (the answer must be in the text)
- Try to avoid questions that elaborate on context or need strong interpretation
- QA systems show text passages as an answer so yes/no questions cannot be answered with “yes” or “no. However, it is possible to get a passage that clearly indicates if the answer is “yes” or “no”. Feel free to ask those questions, just make sure they don’t account for more than 20% of the questions.



**Picture 5: Questions Page**

After having created the standard questions, you can see an overview on the questions page, including the assigned category (Mark 2 Picture 5), the question title and the creation date. In order to change or delete the question, click on the respective button in the “Actions” column (Mark 3 Picture 5).



Picture 6: Creating a standard question



#### Attention:

Each block should not contain more than around 10 standard questions!

## User Management

When the preparations for starting the labelling process are done and you have created all standard questions in different categories, you will need to create accounts for all non-admins/ labellers in your organization. You can do so, by clicking “User Management” in the upper right corner of the tool. When clicking “Create new user” you will be asked to enter the name and email address of the new user. Afterwards, you will be given a password for the new account. Please make sure to save that password and send it to the new user - **it is their log-in password!**

If any of the users loses their password, you have the possibility of resetting their password to a new one, by clicking the small gear sign in the “Actions” column next to the respective user. Please ask your users to carefully save their passwords somewhere, in order to avoid having to reset password continuously.

Now that all labelers have an account and you created all standard questions, **assign the categories/ working blocks to the labelers and clearly tell them to only label their block.**



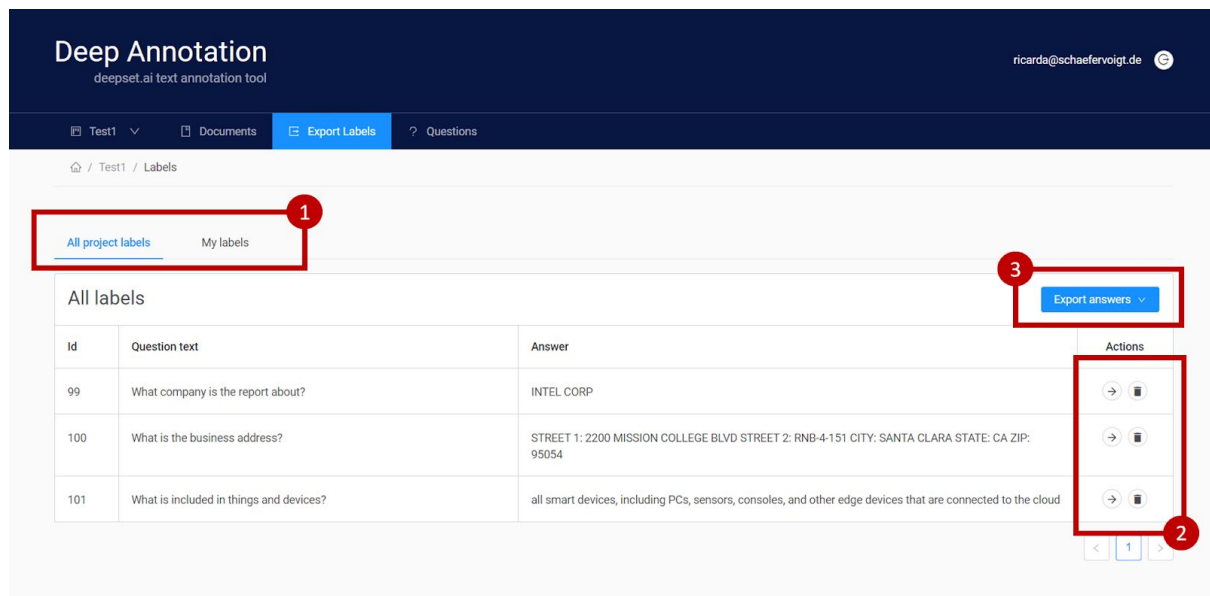
**Attention: The tool is not multi-client capable.** That means if two labellers have the same question for the same document, they will overwrite each others answers. In order to cope with this property, you can either clearly differentiate questions through blocks or you can create multiple projects.

**Let the labelling begin!**



## Exporting labels

When the labelling has been completed for all documents and questions, you can export the labels from the tool. To do so, please click on “Export Labels” in the upper blue bar on the tool’s page.



**Picture 7: Exporting Labels**

On the exporting page for the labels, you can generally choose between exporting only your own or all labels related to the project (Mark 1 Picture 7). In the table, all labels including question and answer text are presented. Via the “Actions” column (Mark 2 Picture 7) you can delete individual labels or get directed directly to them.

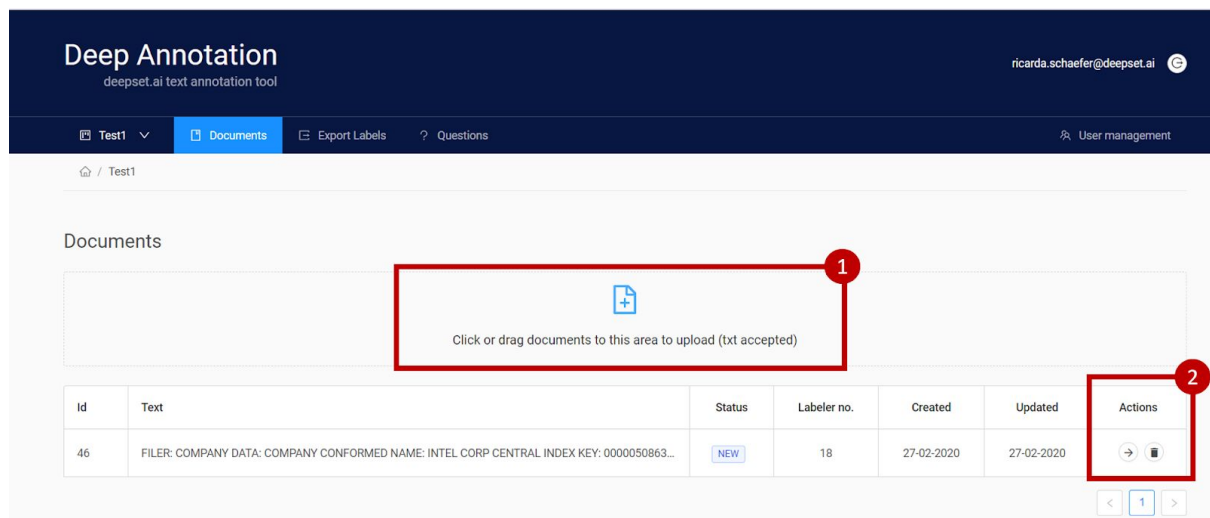
Once you are ready for the export, click on the blue “Export answers” button and select between an excel file or squad format for the export. Your file will then be provided.

## PART 2 - THE LABELLER’S TASK

### Navigating to the right questions

As a labeler, you should received an invitation to the DeepAnnotation tool by your organizations admin, including your email and your password. Enter them on the login page of <https://annotate.deepset.ai/login>.

On the main page, select the project you have been assigned to by hovering over “All Projects” in the upper left corner and clicking on the correct project name. After that, you will get to the document page, where you need to click on the small arrow in the “Actions” column right next to the first document you want to label (Mark 2 Picture 8).



**Picture 8: The Document page**

When you have done so, you will be shown the actual labeling interface and you are almost ready to go. Your admin has assigned you a specific working block/ category. To select the questions of that category click on the assigned letter in the “Questions” block (Mark 1 Picture 9). The questions you need to label will then appear right below (Mark 2 Picture 9).



**Attention: Please only label the questions you have been assigned!** If you label your questions faster than expected, please contact your admin rather than answering questions from other blocks.

## How to label

Look at the category you have been assigned. If the category is anything, but not “U”, please follow these instructions. The general way the process works, is that you need to:

1. **select a question from your question block/ category by clicking on it**, and then
2. **answer it by highlighting the correct words or passages in the document.**

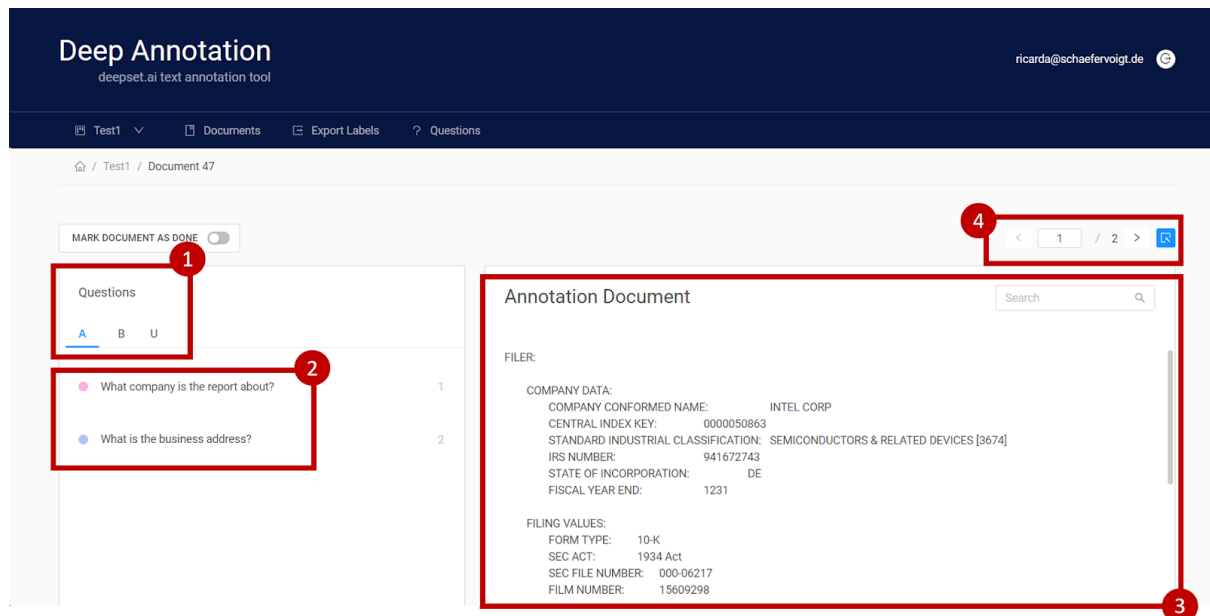
The documents in which you can highlight the answer text is shown on the right side of the page (Mark 3 Picture 9). Each pair of a question and the text you highlighted as an answer to it is considered one label.

**Hint:** You can move between questions either by clicking on them directly or by entering the number shown right next to them.

In case you have misplaced a label you can delete it. Do so by first selecting the respective question and then clicking on the highlighted text passage in the document. You will then be asked if you really want to delete the label.

When you have labelled all questions in your block for the shown document, you will need to move to the next document. You can do that either by clicking the left/ right arrows of your

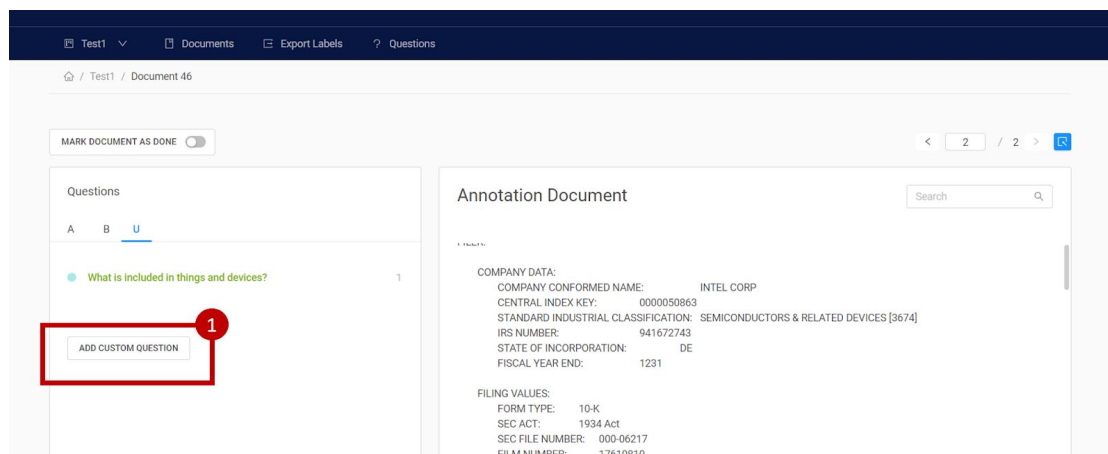
keyboard or by entering the document number/ clicking the little arrow on the upper right corner of the page (Mark 4 Picture 9).



Picture 9: Labelling interface

## Labelling for Category “U”

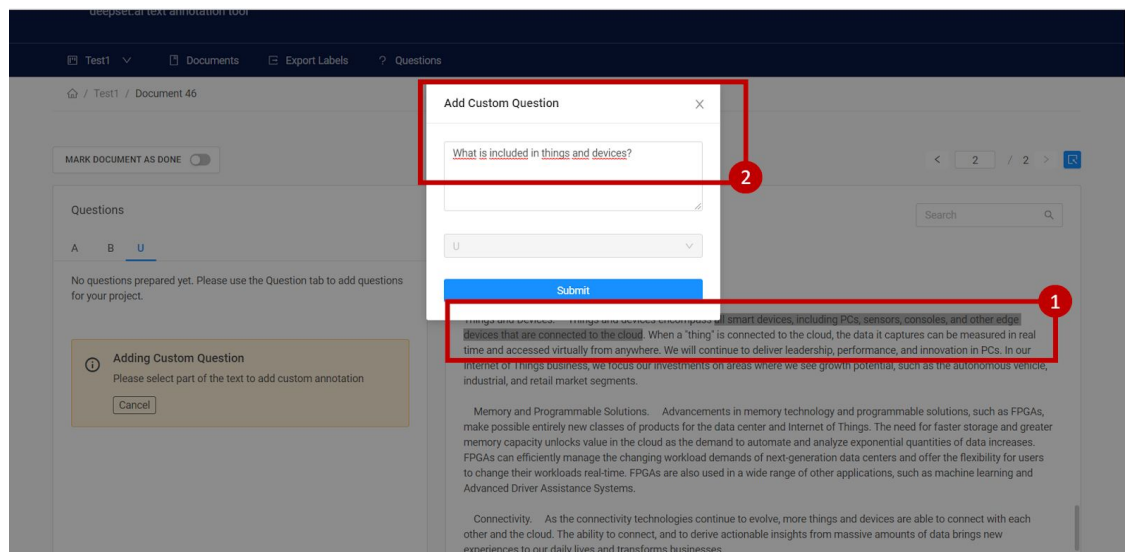
In case you are one of the few people that were assigned to category “U”, your labelling process works a little different than for the rest of the labellers. Rather than selecting pre-defined questions and finding an answer for them, you will highlight text passages and then write a corresponding unique question that fits to the selected answer yourself. Unique in this context refers to the property that this question most likely won’t be applicable to the other documents or text passages.



Picture 10: Category U

When you get to the labelling interface and have selected the category “U”, you see the documents on the right side and the questions on the left side. As you haven’t created any

questions yet when starting, the left side will be empty first. To add a custom question, click the button “Add custom question” (Mark 1 Picture 10). After you have clicked the button, you will be prompted to highlight the text passage (as an answer) that you want to create the corresponding question to. The text will appear coloured, when you have marked it (Mark 1 Picture 11).



**Picture 11: Adding a custom question**

When you have finished the marking, the question input field appears (Mark 2 Picture 11). Enter your question and click “Submit” to finish with the first custom question. After adding a question, it appears on the left side in the “Questions” block. To change or delete a question you write, simply hover over it and click on the pencil or trash symbol at the right side of the question field.

As with all other categories, you can switch between the different documents in order to create questions for all of them.

## General remarks on labelling and question creation

There are some general guidelines to follow while labelling that will ensure consistent, high-quality labels:

### Creating Question:

1. When creating a question, don't use the exact terms from the answer passage, but **think of synonyms or more abstract formulations**
  - *Example*
    - “...**Sales** was 20% up compared to last quarter..” - Question: “How did **turnover** evolve?”, “What was the general movement of the **top-line**?”

2. Don't create question that elaborate on context
  - *Example*
    - **Don't:** "Considering the acquisition of ABC Inc., how did it impact the balance sheet?"
    - **Do:** "How did the acquisition of ABC Inc. impact the balance sheet?" or "Did the acquisition of ABC Inc. impact the balance sheet?"
3. **Yes/No Questions:** Yes/no question cannot be answered by the model with "yes" or "no", but it is possible to get a passage that clearly indicates if the answer is "yes" or "no"
  - *Example*
    - "...as expected, our **sales went up** through introducing a new product for consumers significantly...." - Question: "Was sales impacted by the new product line?"
  - **To consider:** Feel free to ask such questions from time to time, but take care that not more than 1 out of 5 unique questions is such a question

#### Answering Questions:

1. **Number of answers:** for **each question**, there should be **one single, connected passage** per document
  - in case you see multiple answers within one document, use the unique question mode
2. **Length of answer: focus** on a passage that **answers** the questions **to the point** and do **not mark whole abstracts**
  - *Example*
    - **Don't:** "Does the company envision 3D sensing to become the majority of sales? - *Answer:* "Well, certainly 3D sensing is going to be an important portion of our business. However, we still see the other part -- the power amplifier for 5G has also, as we presented earlier, also has a very high growth rate in the following years. So we believe our PA business and the infrastructure business will continue to have a significant portion of our total business. Of course 3D sensing is also a big part of our business, as well, but probably it would not be the majority one."
    - **Do:** "Does the company envision 3D sensing to become the majority of sales? - *Answer:* "...3D sensing is also a big part of our business, as well, but probably it would not be the majority one..."
3. **Bulleted lists:** this is feasible, so you can mark them when you think it is suitable (consider the remarks to "Length of answer")