

Questions marked with a \* are required

---

## Introduction - Active machine learning and the annotation of text

This survey aims at investigating the extent to which Active Learning (AL) is used in the context of Natural Language Processing (NLP), as well as addressing the reasons to why (or why not) AL has been found applicable to a specific task.

Supervised machine learning methods have been successfully applied to many NLP tasks in the last few decades. While these techniques have shown to work well, they require large amounts of labeled training data in order to achieve high performance.

Such training material, however, is typically not available for specific domains or problems. Thus, one could say that the annotation of data is becoming a major bottleneck in the development of NLP systems.

AL addresses this problem. The main idea is to give the control of which data points need to be annotated to the learning mechanism in order to achieve a high model performance with as little labeled examples as possible.

This survey targets participants who were or are currently involved in the annotation of training material for any kind of NLP task. It is not required that you are familiar with or have used AL. It will take approximately 10 minutes to fill out the survey. Results will be published online and sent by email to the participants (if desired).

The survey is completely for non-commercial, academic purposes. Any user-specific data will be kept in confidence and not passed to a third party. The survey is initiated by Katrin Tomanek (Jena University, Germany) and Fredrik Olsson (SICS, Sweden). If you have any question, please feel free to contact us.

So, lets start...

Have you ever used Active Learning in order to speed up annotation/labeling work of any linguistic data? \*

Yes

No

## Part I - Questions on annotation project supported by Active Learning

The following questions relate to type and scope of previous annotation projects where you have used AL. In case you performed more than one such annotation project, please answer the following question with respect to **one** project only.

Note: we will refer to the single documents, words, sentences etc. which had to be labeled as (*annotation*) instances.

Which of the following is the annotation project (where you used AL) you are reporting on? (Select one alternative) \*

Most recent one

Most successful one

Largest one in terms of labeled units

Largest one in terms of spent annotation time

Most interesting one

Other

What kind of annotation had to be done? (Multiple answers allowed) \*

Document classification

Part-of-speech

Chunk

- Parse
- Word senses
- Sentiment
- Named entity
- Relations/Events
- Other

From which domain(s) were the annotated instances? (Multiple answers allowed) \*

- Newswire
- Bio-medical
- Other

What was the principal language? (Select one alternative) \*

- English
- Hindi
- Mandarin
- Arabic
- Spanish
- French
- German
- Other

How many person months were spent on annotation? \*

How many annotators have been involved? \*

What was the size of the resulting, labeled corpus in terms of annotated instances (please provide the annotation instance type in brackets, e.g., "1000 (tokens)") ? \*

How did you assemble the corpus of unlabeled annotation instances (i.e., the pool from which your AL mechanism selected)? (Select one alternative) \*

- We took all annotation instances available to us
- Random subsample of all annotation available instances
- Subsample of all annotation available instances based on introspection
- Subsample of all annotation available instances based on keywords
- Other

What was the size of this unlabeled corpus in terms of annotation instances (please provide the annotation instance type in brackets, e.g., "1000 (tokens)")? \*

## Part II - Questions on applied Active Learning techniques and experiences

What kind(s) of active learning approach have you used? (Multiple alternatives allowed) \*

- Uncertainty sampling
- Query-by-Committee (including variants such as Query-by-Bagging, Query-by-Boosting, Co-testing etc)
- Variance Reduction
- Expected error reduction
- Other

What was the base learning algorithm used (e.g., decision trees, Maximum Entropy)? \*

How did you decide when to stop the annotation process? (Select one alternative) \*

- Evaluation of model learned on annotations with a held-out gold standard
- Money and/or time available for annotation used up
- All relevant documents annotated
- Expected gains in model performance compared to additional annotation costs below critical threshold
- Other

What was your primary motivation to use AL? (Select one alternative) \*

- Minimize labeling cost
- Gather experience with/test AL
- Test your annotation guidelines against hard cases
- Other

Have your expectations with respect to AL been met? (Select one alternative) \*

- Yes
- No
- Partially

If not, please describe the problems/shortcomings.

Please estimate the reduction of annotation effort in terms of *annotation units in percent* (leave blank if you cannot give an estimate).

Please estimate the reduction of annotation effort in terms of *time saved in percent* (leave blank if you cannot give an estimate).

Please describe any other experiences regarding your practical use of AL

Would you use AL again in an annotation project of this type and scope? (Select one alternative) \*

Yes

No, because

## Part I - Questions on previous annotation projects

The following questions relate to type and scope of your previous annotation projects. In case you performed more than one such annotation project, please answer the following question with respect to **one** project only.

Note: we will refer to the single documents, words, sentences etc. which had to be labeled as (*annotation*) instances.

Which of the following is the annotation project you are reporting on? (Select one alternative) \*

Most recent one

Most successful one

Largest one in terms of labeled units

Largest one in terms of spent annotation time

Most interesting one

Other

What kind of annotation had to be done? (Multiple alternatives allowed) \*

Document classification

Part-of-speech

Chunk

Parse

Word senses

Sentiment

Named entity

Relations/Events

Other

From which domain(s) where the annotated instances? (Multiple alternatives allowed) \*

Newswire  
 Bio-medical  
 Other

What was the principal language of the annotated texts? (Select one alternative) \*

English  
 Hindi  
 Mandarin  
 Arabic  
 Spanish  
 French  
 German  
 Other

How many person months were spent on annotation? \*

How many annotators have been involved? \*

What was the size of the resulting, labeled corpus in terms of annotated instances (please provide the annotation instance type in brackets, e.g., "1000 (tokens)")? \*

How did you assemble the corpus of annotation instances to be labeled? (Select one alternative) \*

We took all annotation instances available to us  
 Random subsample of all available annotation instances  
 Subsample of all available annotation instances based on introspection  
 Subsample of all available annotation instances based on keywords  
 Other

How did you decide when to stop the annotation process? (Select one alternative) \*

Evaluation of model learned on annotations with a held-out gold standard  
 Cross-validation of model on annotations  
 Money and/or time available for annotation used up  
 All relevant documents annotated  
 Expected gains in model performance compared to additional annotation costs below critical threshold

Other

## Part II - Questions on alternative annotation instance selection

Are there specific reasons to why you did not apply Active Learning in this project? (Multiple alternatives allowed) \*

- Never heard of Active Learning
- Insufficient expertise/knowledge
- Did not want to spend overhead in implementing Active Learning-based annotation environment
- Did not meet my project's specific requirements
- Wanted a corpus with a different instance distribution
- Was not convinced that Active Learning would reduce labeling cost
- Doubts that Active Learning would work well in my scenario
- Other

Please elaborate on the following two aspects if they apply to you: Why did Active Learning not fit your specific requirements? What were your doubts concerning the applicability of Active Learning to your scenario?

Would you consider applying active learning in future annotation projects? (Select one alternative) \*

- Yes
- No
- Maybe
- Only under specific circumstances:

## Part III - Background Information

Where are you currently working? (Select one alternative) \*

- Academia
- Industry
- Governmental organization
- Other

What is your educational background? (Select one alternative) \*

- Linguistics
- Computational linguistics
- Computer Science/Informatics
- Mathematics

Other

Which field are you currently working in? (Select one alternative) \*

General Text Mining

Bio Text Mining

Machine Learning

Other

Do you have any comments on the survey?

Up until this point, the survey has been completely anonymous. Answering the following questions will make you known to us, and therefore the section below is optional.

You have been reporting on the type and scope of one of your annotation projects. What was the name of the annotation project? Are there publicly available references such as URL:s?

Please enter your email address if you want to be informed about survey results

May we contact you with follow-up questions? (requires that you have specified your email address above)

Yes

No

Please contact [katrin.tomanek@uni-jena.de](mailto:katrin.tomanek@uni-jena.de) if you have any questions regarding this survey.

Powered By:  
**QuestionPro** 

[Survey Software](#) | [Idea Management](#) | [Polls](#)