

## Lab 5: Aggregation Framework

We will be using the same Twitter data base that we used in Lab 4 about Map/Reduce.

### Quiz Time: Twitter

Map/Reduce is the big hammer when it comes to data aggregation with MongoDB. But since Version 2.2 we have another option the Aggregation Framework which has limited functionality compared to Map/Reduce but which is just enough for about 80% of the typical use cases For example you can get an alphabetically sorted list of all German Twitter Users as following:

```
> db.tweets.aggregate(  
  { $project : { _id : 0, name : "$user.name", language : "$user.lang" } },  
  { $match : { language : "de" } },  
  { $sort : { name : 1 } }  
);
```

In a similar way we want to solve some challenges using the Aggregation Framework.

### Question 1: Languages of tweets

Which is the second mostly used language in the tweets and how many tweets are there in this language?

### Question 2: Most popular Twitter users

What are the names of the Top-3-Twitter Users measured by the number of followers

### Question 3: Verified users with 4-digit follower count

How big is the number of verified users with at least a 4-digit count of followers (meaning > 999)?

## Question 4: Inactive Twitter Users

If a user has no followers and no friends, one can assume that the user is inactive. Using these criteria, how big is the number of inactive Twitter users?

## Question 5: Most popular links

In the array `entities.url` there is one entry per link posted by the user in the text of the tweet (meaning one link per field `entities.url.url`). Using these URL's one can create a ranking of the most popular links on Twitter. What is the Top-10 of the most popular links?