

750 Volunteers Transcribing 8.5 Million Records Online — an Evaluation

Dr. Jesper Zedlitz

Research Group for Communication Systems

Department of Computer Science

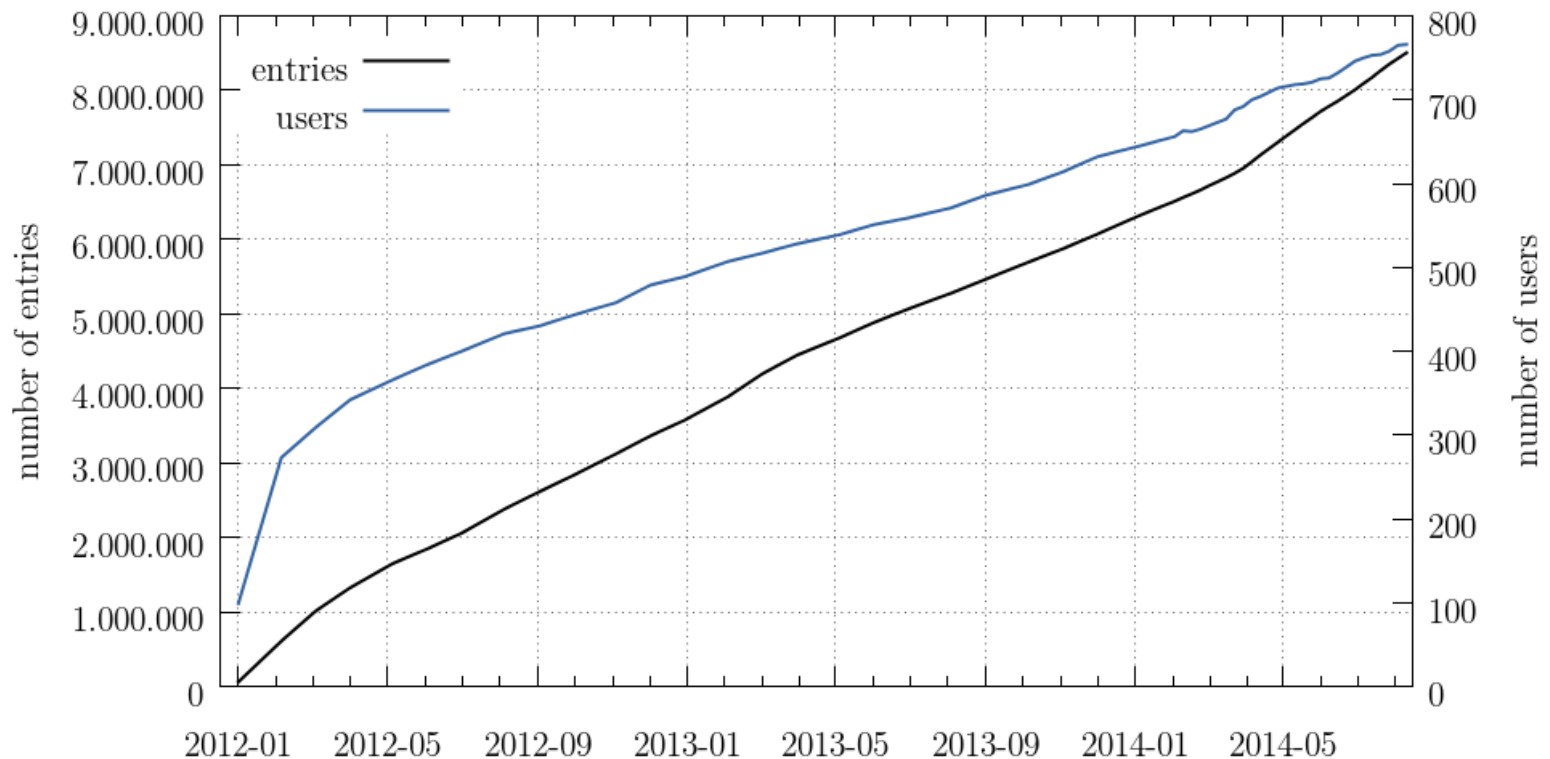
Christian-Albrechts-University in Kiel

Project Topic

- German 1st World War casualty lists
 - one of the most important preserved sources for German WW1 soldiers
 - published between 1914 and 1919
 - 2.535 volumes
 - newspaper format (305 x 455 mm)
 - Gothic print
 - more than 31.000 pages

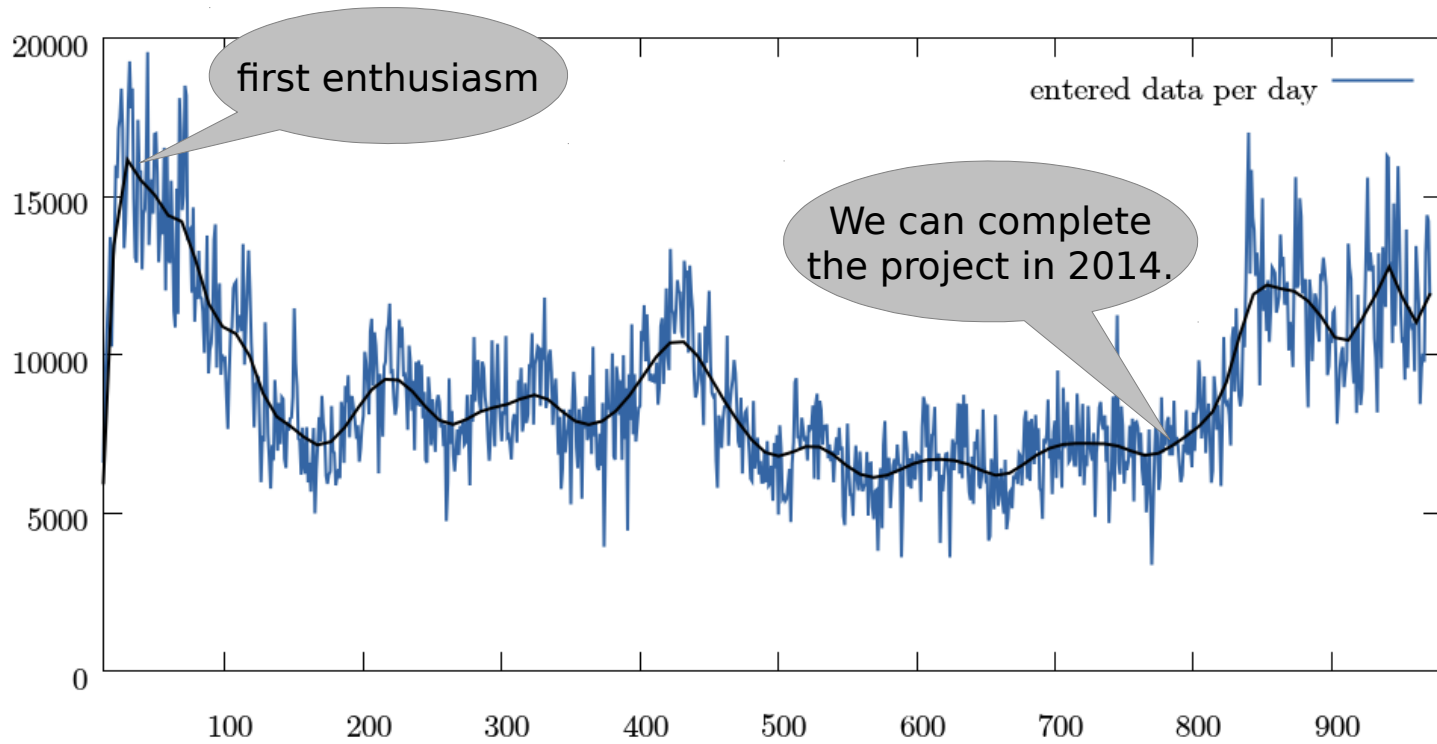


- Entered records and registered users



Experience with Crowdsourcing

- Entered entries per day



- H1 80% of the data is entered by 20% of the users. (Pareto Principle)
- H2 Most work is done in the evenings and weekends.
- H3 „Power users“ stay longer in the project.
- H4 Number of entries per day increases sharply at entry. Drops to a long-term average.
- H5 Typo rate decreases with experience.
- H6 „Power users“ make fewer mistakes.
- H7 Users who leave quickly, contribute little and cause a lot of work.