Improving NER Ground Truth with Crowds

OANA INEL, LORA AROYO - Vrije Universiteit Amsterdam
{oana.inel,lora.aroyo}@vu.nl

WHAT'S WRONG WITH GROUND TRUTH DATASETS?

experts make mistakes

[One of the them] was an eminent scholar at Berkeley.

Albert Lutuli was visited by United States Senator Robert F. Kennedy, who was visiting **[South Africa]**. -> ORGANIZATION

experts are inconsistent ROLE&PERSON

After a long decline, Bologna was reborn in the 5th century under **[Bishop Petronius]**.

William Golding was appointed Knight Bachelor by the **[Queen]** **[Elizabeth II]** in 1988.

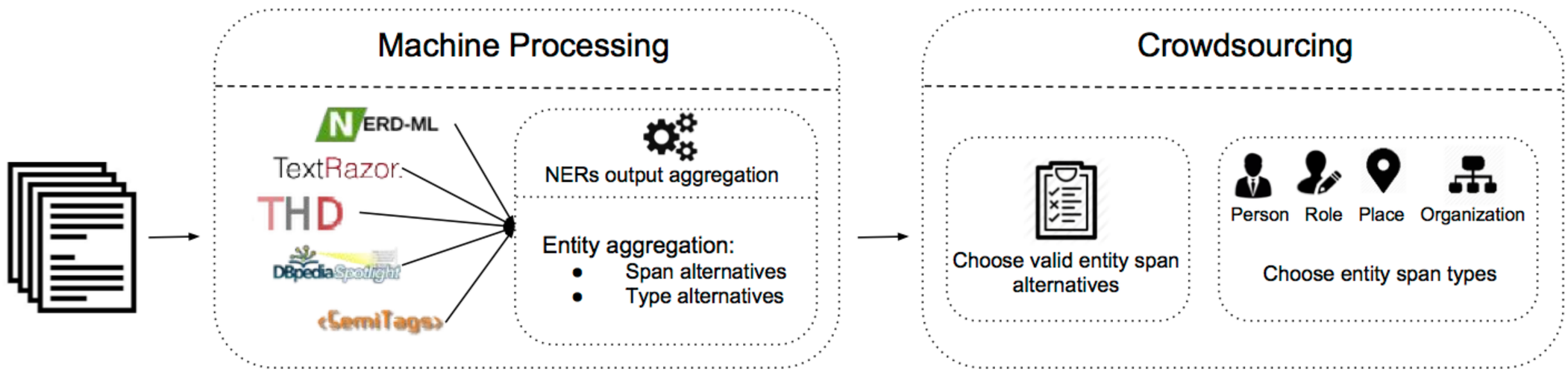
experts are inconsistent LOCATION

… worked at the Basel Institute for Immunology in **[Basel, Switzerland]**.

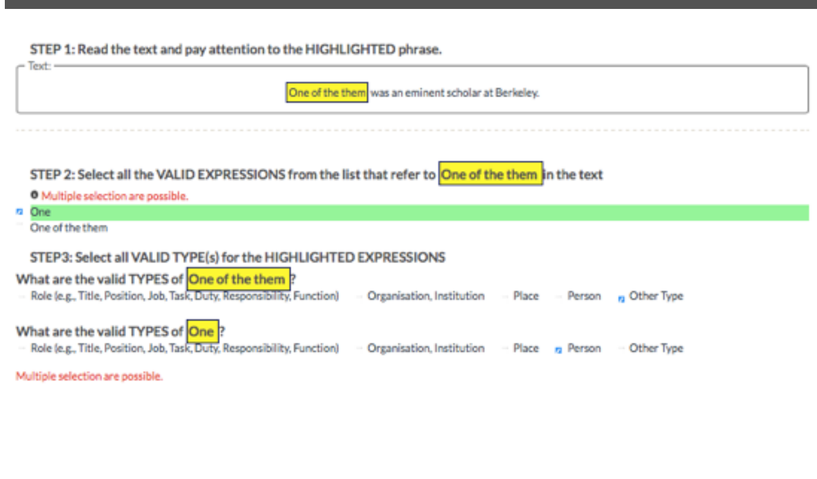
Alfred Nobel was born on 21 October 1833 in **[Stockholm]**, **[Sweden]**, into a family of engineers.

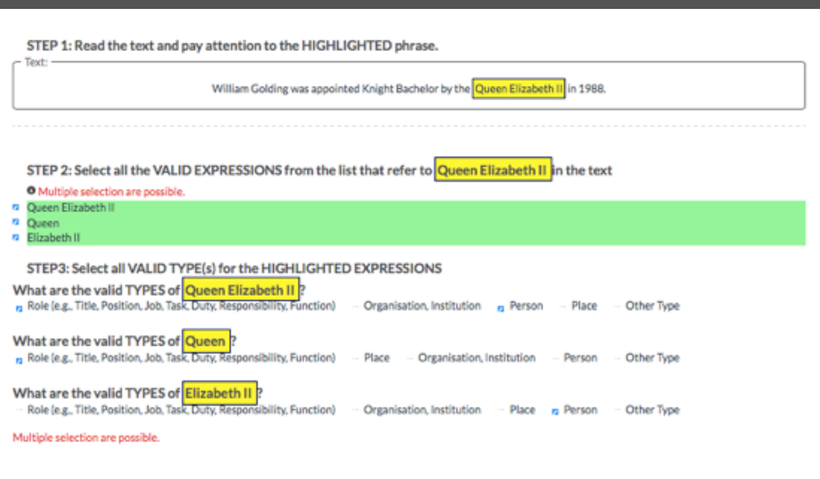
How can crowds help?

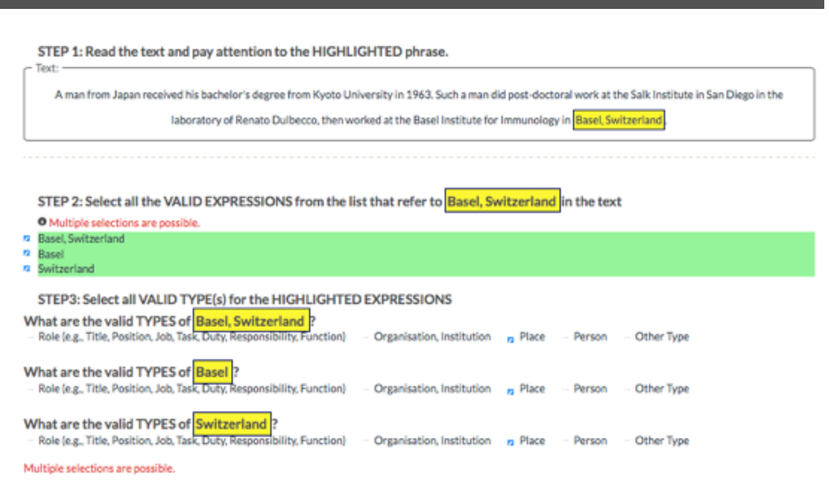
HYBRID MULTI-MACHINE CROWD APPROACH



CROWDSOURCING EXPERIMENTS

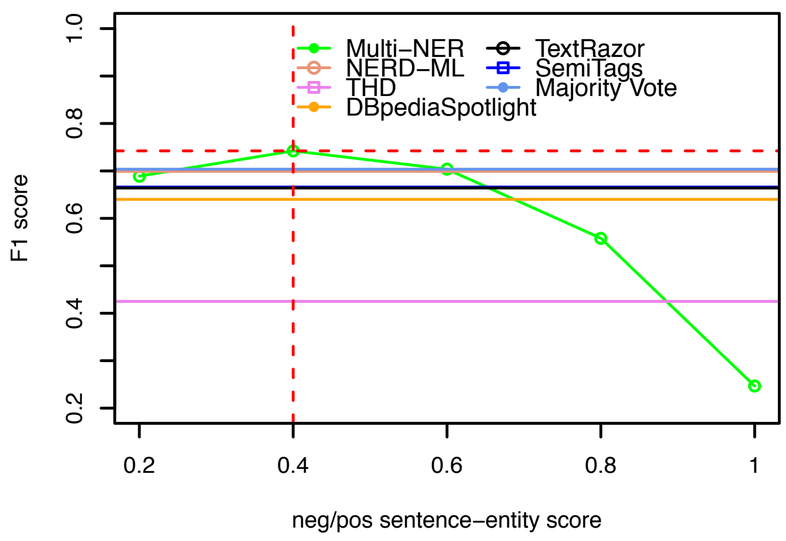






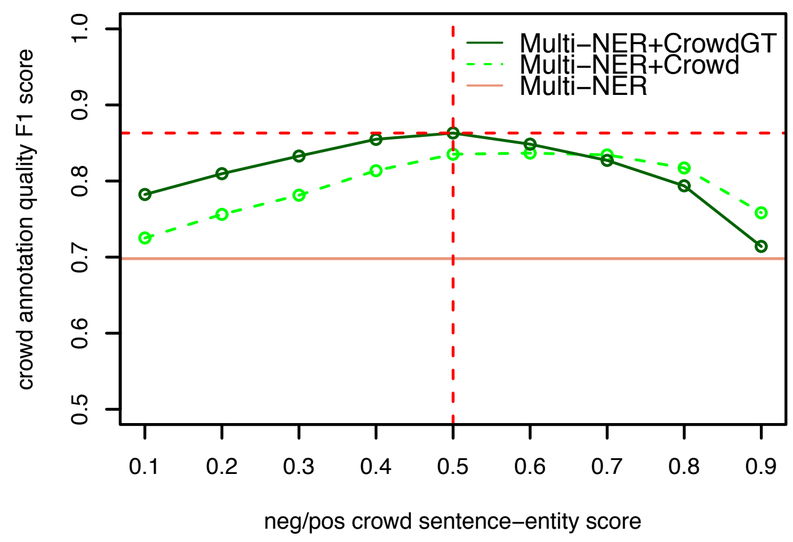
RESULTS AND CONCLUSIONS

Annotation quality F1-score per negative/positive sentence-entity threshold



At a sentence-entity score threshold >= 0.4 **MultiNER** outperforms all **state-of-the-art NER tools** (statistically significant, $p < 2.2e^{-16}$) and the **majority vote** approach (sentence-entity score threshold = 0.6, statistically significant)

Annotation quality F1-score per negative/positive crowd sentence-entity threshold



Multi-NER+Crowd outperforms **Multi-NER** for each crowd-entity score threshold (statistically significant, $p = 8.128e^{-11}$). The crowd is able to identify issues in **GT** and correct them (**Multi-NER+CrowdGT -** statistically significant)

the crowd is **diverse** and captures a multitude of **perspectives** and **granularities**
the crowd is **more consistent** than the experts
the crowd provides a **more reliable GT** by dealing with the intrinsic **ambiguity** of natural language

Website crowdtruth.org

Data Publishing Page data.crowdtruth.org/crowdsourcing-ne-goldstandards/

Code Repository github.com/CrowdTruth

VU VRIJE UNIVERSITEIT AMSTERDAM

#crowdtruth
@laroyo, @oana.inel

CrowdTruth
The framework for crowdsourcing ground truth data