



Exploration of Crowdsourcing and Information Retrieval Techniques for Early Detection in the Area of Food Safety

Note: This is a Report Summary. The full report is attached to this document.

*Abraham Bernstein, Luca Rossetto, Cristina Sarasua and Kathrin Wardatzky
Department of Informatics, University of Zurich, Binzmühlestrasse 14, 8050 Zurich*

Key words

Early Detection, Crowdsourcing, Text Classification, Collaborative Decision Making

Aim of the study

The project “Exploration of Crowdsourcing and Information Retrieval Techniques for Early Detection in the Area of Food Safety” (EXPLO) aimed to identify and analyze requirements, as well as investigate how hybrid human-machine methods could help the FSVO in analyzing and labelling different types of information.

As stated in the initial proposal, the expected deliverable of this pilot project was “a report describing observations, requirements and the design of potential solutions to improve the current workflow for early detection of food safety issues using crowd- and machine intelligence.” In order to attain this result, the project was conducted in two stages: In the first stage, we conducted a series of interviews in order to understand the characteristics, limitations and potential of the current early detection process. In the second stage, we conducted a survey-based experiment in order to assess the feasibility of multi-annotator hybrid systems.

Material and methods

The first stage started with a “brain-dump” meeting with the process owner (PO) Thomas Lüthi and intern Cornelia Wagner. It then led to a set of semi-structured interviews with interview partners chosen from a pre-selected list of people that was compiled by the PO.

Given the findings obtained during the interviews, we inferred that many of the challenges arising from the existing process are connected with the current collective decision-making process. To investigate possibly better alternatives, we conducted a survey with experimental elements in order to study the relevance assessment process followed by various types of expert users involved in the early detection scenario. The survey was conducted using an online survey platform, where subjects were confronted with two documents. They were then asked what information about the documents they would need to make an assessment about their relevance, the documents’ relevance, and the confidence of their assessment.

Results and significance

The insights found in the interviews were classified into 14 categories that can be clustered into the two groups (i) information to understand the process and how the stages are connected and (ii) the assessment of the current early detection approach by the interviewees. The mentioned challenges were then categorized into three main challenge areas: the current early detection process in general, the external image of the early detection process, and aspects related to the humans that are involved in the process. The latter challenge area contains sub-categories in which the participants described challenges that are related to the information searching and the decision-making processes.

The survey conducted in the second stage indicated that whilst the subjects showed a low inter-rater agreement their collective judgement had a high conformance with the assessment by the process owner. It also highlighted the kinds of information that the subjects relied on for their judgement.

The findings indicate that it should be possible to generate document relevance assessments using a multi-annotator process. Due to the small sample size, it is currently, however, unfeasible to make any statements about the most suited aggregation method of these individual assessments or whom to involve to make the most effective use of individual skills.

Please read the attached full report for a complete set of results, limitations, and recommendations.

Publications, posters and presentations

None. This was an exploratory Pilot.

Project 4.20.06 (Bezeichnung RB 20EC)

Project Duration 26 June 2020 – 15 December 2020