Extraction Form Description

# Definition of a Detection Method

A detection method to be considered for us is a set of steps that can be reproduce to detect fakes/ spam / fakers or spammers in the context of online product reviews or online service reviews. It is important that we only consider methods that are described in a sufficient way to provide information about how the method works and can be used. It is not sufficient to be considered for us as a fake detection method if it is claimed to have a method executed or to be able to detect fakes / spam.

# Hints for extracting in general

We extract only information from the papers where we finally decided to include them in our SLR.

For extracting the data read the paper and try to fill out the fields of the method table with the information provided in the paper. Do not add your own guesses. It is best to quote the paper directly to not add any interpretation from yourself. Put all quotes inside quotation marks. If you have multiple quotes within one field, mark them properly. If you shorten quotes in-between insert […] to indicate that you have reduced something. If you have formulas containing sub scripts indicate the subscript start by using a “\_”. Formulas containing superscripts should indicate the superscript start by using a “^”. Use appropriate parenthesis to ensure that the formula can be understood in Excel.

# Procedure for Extracting Data

We are extracting the methods being presented in primary studies into the method table.

Secondary Studies are left out first and get an annotation in the comment column for the paper.

Secondary Studies as well as Elements in the related work are considered for Phase II.

If you are going to extract the methods being proposed in a paper, please mark in Paper table “Methods extracted” with yes. If the paper does not include a primary method, mark the paper with a “Yes, but no primary methods”.

If a paper proses multiple methods you should extract them as multiple methods in the methods table.

## Quality Crieria for our Extraction

Please use the following quality assessment criteria within the first phase:

|  |  |
| --- | --- |
| Criteria | Description |
| Pri | Paper contains a primary research. We require that papers have to present a new or update method |
| MethodDec | The method description must be written in a way that readers are able to understand the purpose of the method, how it works as well as the major steps for the method. |
| MethodEval | The method’s evaluation should be presented in an understandable way. Readers have to get an understanding about the data being used for evaluation and how the evaluation was performed, as well as the results of the evaluation. |

The Pri criteria is a requirement to continue with the extraction. The MethodDec criterion is a requirement to continue with the extraction. If the MethodEval criterion is not fulfilled, we extract the method, but not the evaluation.

If any of the criterion is not fulfilled the impact for our SLR is lowered. This lower impact should be noted in Table “Paper” column “Impact” with a low. The reasons for lowering the impact should be noted in the “Impact Reasons column”.

**The decision that a paper is not fulfilling one or more quality criteria has to be taken by two different people.**

**These criteria should act as a minimum requitement for the papers to be considered. Fullfilling them does not imply that a paper has a good quality.**

# Description for the Extraction Fields for “Method Table”

* **MethodID**: Number of the Method you have extracted. Just auto increment the value by 1 for each method.
* **Name of the Method**: Copy the name of the method or approach as described in the Paper.
* **PaperID**: Number of the Paper from the paper table, where the method was found
* **View**: Either “Review” or “Reviewer”. We use the “View” field to distinguish between detecting spam and detecting spammer. The “review view” is focused on the performing the analysis based on reviews, the “reviewer view” is focused on performing the analysis based on data and activities about the reviewer itself.
* **Level**: One of the following “Review”, “Product” or “Source”. This field is used to distinguish the level of information being used. The methods being on “Review Level” are performing their analysis just by using a review. The ones being on “Product Level” consider multiple reviews of the same product and / or information about the product itself when executing the method. Methods on “Source Level” consider in addition characteristics or features of the data source itself and might use multiple products to detect spam.
* **Method Assumptions**: If the method builds on certain assumptions on the data, spam or spammer list that here. These assumptions might influence how well a method fits into the real world and how well it could be transferred to other data sources or products. If you do not find assumptions please add an “-“.
* **Domain**: The domain of products or services where the method is designed to be applicable or is explored so far. If the domain is not clear please add an “-“. Domains could be for example “products”, “services”, “travel”, “apps” etc.
* **Method** **Description**: Extract a detailed description about the method including purpose, goals, context as well as data being used. The description might also contain details about the methods implementation.
* **Method** **Steps**: List the steps as described in the paper to execute the method. This should be a high-level algorithm.
* **Evaluation** **Description**: Extract a detailed description about the evaluation. This includes the data set being used for evaluating it as well as the experimental setup.
* **Evaluation** **Results**: List the results of the evaluation, including strengths and weaknesses as well as a conclusion how applicable the method was.
* **Language**: The language for which the method was designed to detect spam or spammers. If the language source is not names in the paper please add an “-“.
* **Datasource**: The data source with product reviews being used for the method. If the data source is not names please add an “-“.
* **Degree** **of** **automation**: Describe how well the method is automated. This aspect is only focused on the fake detection method. It does not consider the degree of automation how to get the reviews or other things. This is either “not”, if it has to be fully performed by humans; “semi”, if the method is partially automated and partially manual; “full”, if the method works fully automatically. If the paper does not give insights on that matter, please leave the field out.
* **Comment**: Used to add information worth to be extracted or noted, which is not fitting in any field so far.