

# Accurate Retrieval of Corporate Reputation from Online Media Using Machine Learning

Achim Klein  
Information Systems 2, University  
of Hohenheim, 70599 Stuttgart,  
Germany  
Email: achim.klein@uni-  
hohenheim.de

Martin Riekert  
Information Systems 2, University  
of Hohenheim, 70599 Stuttgart,  
Germany  
Email: martin.riekert@uni-  
hohenheim.de

Velizar Dinev  
Information Systems 2, University  
of Hohenheim, 70599 Stuttgart,  
Germany  
Email: velizar.dinev@gmail.com

**Abstract**—Corporate reputation is an economic asset and its accurate measurement is of increasing interest in practice and science. This measurement task is difficult because reputation depends on numerous factors and stakeholders. Traditional measurement approaches have focused on human ratings and surveys, which are costly, can be conducted only infrequently and emphasize financial aspects of a corporation. Nowadays, online media with comments related to products, services, and corporations provides an abundant source for measuring reputation more comprehensively. Against this backdrop, we propose an information retrieval approach to automatically collect reputation-related text content from online media and analyze this content by machine learning-based sentiment analysis. We contribute an ontology for identifying corporations and a unique dataset of online media texts labelled by corporations' reputation. Our approach achieves an overall accuracy of 84.4%. Our results help corporations to quickly identify their reputation from online media at low cost.

## I. INTRODUCTION

A great variety of firms offer an even greater variety of products and services to consumers and other businesses and strive to build up a strong corporate reputation. Corporate reputation can be defined as the collective perception and judgment of the sentiment (i.e., feeling, opinion) about a corporation and its products or services by its stakeholders. Reputation as a necessary condition for differentiation and corporate success has become one of the central themes in all its facets for both practitioners and the scientific community [1]. The ability to quickly assess current movements in the own and the competitors' corporate reputation is crucial for operative decision making, corporate planning and strategy as well as for external investment decisions.

The important role of corporate reputation has been confirmed through extensive research. Shefrin and Statman show that corporations with good reputation represent good long-term investment opportunities [2]. These corporations with a good corporate reputation are more likely to receive funding on the capital markets at better conditions. The positive relationship between corporate reputation and investor expectations about a firm has been supported again later by Shefrin [3] and MacGregor et al. [4] pointing

toward a stable relationship. Corporate reputation is shown to be positively related to return on sales and assets, sales, earnings per share, price-to earnings ratio, dividend yield, net income of a company, and customer loyalty [5-8].

Studies in the field of corporate reputation [5-8] have in common that authors use either the Fortune magazine's reputation index published in the annual survey "Most admired companies" or conduct a survey on their own to measure corporate reputation. The Fortune magazine's survey is conducted annually among more than 8000 managers and financial analysts. It rates around 700 companies according to their innovativeness, people management, use of corporate assets, social responsibility, global competitiveness, quality of management, financial soundness, value as a long-term investment and product quality or service quality.

The use of surveys for measuring corporate reputation should be assessed critically because it does not cover all stakeholder groups of a company. It has been shown that the resulting reputation ratings reflect mostly the perception of the financial perspectives of a company [9], [10]. Thus, the meaningfulness of such ratings is limited. Furthermore, the low update frequency of the reputation index and the limitation to 700 companies reduces its usability further. Conducting an own survey is costly, time consuming and often covers stakeholders only partly (e.g., [3], [8]).

Nowadays, online media represent a very good source for reputation related comments by customers of companies. However, measuring corporate reputation from online media is a dynamic and challenging problem. The Internet in general extends the reach, speed and intensity of news [11]. There is a great number of online media outlets where people express their opinions about corporations and their products. Because of the volumes of textual data, manual processing is practically impossible. Furthermore, numerous factors that influence reputation need to be considered. However, an automatic retrieval approach using textual content from online media would be an efficient and holistic way to measure corporate reputation.

We propose to combine an information retrieval approach with sentiment analysis methods for automatically analyzing corporate reputation in online media. We contribute an ontology for identifying corporations in the first place. For

analyzing corporate reputation in text, we contribute a unique dataset of human annotated reputation texts. We use the dataset for corporation-specific reputational sentiment analysis using a machine learning classifier. Our work helps corporations to efficiently measure reputation, which is an important factor for the performance of a corporation.

The remainder of this paper is organized as follows. In section 2, corporate reputation is defined and approaches for measuring corporate reputation are presented. Section 3 specifies the problem. In section 4, the proposed reputation measurement approach is described. In section 5, we present our dataset of annotated reputation texts and evaluate our reputational sentiment classifier. Section 6 concludes.

## II. RELATED WORK

### A. Defining Corporate Reputation

Corporate reputation has been a popular topic in different streams of research, leading to a large amount of definitions of corporate reputation (e.g. [12-16]). According to [13,14] we do not use the terms corporate reputation, image and identity interchangeably. Based on [12-16] we define corporate reputation as the collective perception and judgment of the sentiment (i.e., feeling, opinion) about a corporation and its products or services by its stakeholders. Corporate reputation can be positive or negative [17]. Corporate reputation arises from the ability of a corporation to uphold social and institutional norms and values and to satisfy the needs and desires of its stakeholders. Corporate reputation forms through the appealing “character” [15] of a corporation and in the comparison with other entities.

### B. Approaches for Measuring Corporate Reputation

Most of the empirical reputation research uses the Fortune’s magazine “Most Admired Companies” (FMAC) index for measuring reputation [18]. It is based on a survey of senior executives and directors conducted annually. Companies with revenue of at least 10 billion \$ and at least the 15-th biggest revenue in their industry are ranked according to 9 “attributes of reputation”. The use of Fortune’s reputation data is rightfully criticized because it was shown to mostly reflect only the financial performance of a corporation [9], [10]. Surveying only senior executives and directors neglects all other stakeholder groups. The FMAC index also suffers of industry effects because the surveyed managers are explicitly asked to rate the corporations in comparison only to the other corporations in a particular industry [19]. The limited availability and frequency of reputation data (i.e., the reputation index refers to only the largest corporations) further limits the use of Fortune’s index for operative decision making.

“Britain’s most admired companies” (BMAC) of Management Today offers another publicly available reputation index. It is structurally very similar to the FMAC [20]. Similarly to FMAC, managers rate companies according to nine [20]. The critique to FMAC largely applies to BMAC as well because of the similarities between the two surveys.

“Reputation Quotient” (RQ) is a reputation ranking of the 60 “most visible” companies in the U.S. [21]. The companies are rated on 20 attributes distributed over six components of corporate reputation [21]. 22480 randomly selected respondents’ rated one or two companies. Each company is rated by at least 279 people. RQ is theoretically more founded than FMAC/BMAC but its commercial orientation complicates a closer examination. The fact that only the 60 most popular companies *at the time* are rated limits the usability both for research and practice because of resulting gaps in the time series and the small amount of observations.

Reputation can be also measured by conducting an own survey. This technique was employed by [3], [8], [22-24]. Modifications of the classical written (online) surveys like Verbal Protocol Analysis (taping, coding and analyzing the answers of respondents) [23] and the use of personification metaphor (rating a corporation on a five-point scale in regard to 42 items that load onto five orthogonal character factors) [25] have also been proposed.

### C. Research Gap

The reviewed studies on corporate reputation measurement have one major flaw: they do not cover all relevant stakeholder groups. This fact draws attention to the difficulty of conducting a representative survey of corporate reputation: it is very costly and time consuming. Conducting such a survey on a regular basis and for many corporations is practically impossible for smaller corporations.

We propose a different approach to measure corporate reputation in an automatic, efficient, and more holistic way by retrieving corporate reputation-related textual content from online media and using a sentiment analysis approach.

## III. PROBLEM SPECIFICATION

A document from online media may express reputational sentiments on multiple corporations [26]. Sentiments referring to multiple corporations can have different sentiment polarities. The problem is to classify the reputation sentiment polarity contained in a document with respect to each sentiment object separately. By classifying reputation sentiment, all factors influencing reputation should be considered.

## IV. APPROACH

This section describes our approach for extracting corporate reputation from online media texts. A machine learning based classifier is used for reputational sentiment classification. This approach does not require costly and time-consuming optimization of a knowledge base [29]. It is computationally efficient due to the linear classifier [30].

First, each document was pre-processed with natural language processing techniques, similarly to [27]. The preprocessing includes tokenization, sentence splitting, part of speech (POS)-tagging, and morphological analysis for lemmatization. Following [27, 28] the pre-processing was

implemented by GATE’s information extraction system [31].

The pre-processing includes ontology-based entity recognition. For this purpose, an ontology was developed based on [27]. The ontology contains all corporations from the Dow Jones Industrial Average, S&P 500 and S&P 600 indices, and various European and US banks that are also present in our reputation text dataset. For each corporation, hand-curated labels were defined for textual identification.

Second, we extracted relevant text segments that refer to a certain corporation, which were identified by the ontology. Following [32], the relevant text segment is defined as 25 words either side of the mention of a corporation. Then, all text segments referring to the same corporation within one article are concatenated.

Third, a linear kernel soft-margin Support Vector Machine (SVM) is applied successively on each of these text sections, as it has been shown to perform text classification tasks on state-of-the-art level, when given limited training data [30], [33]. We used the default hyperparameter configuration of SVM. The hyperparameter of the SVM for the costs associated with allowing training errors was set to 1. The features used by SVM are frequency counts of unigrams in a document [34]. The feature space of SVM contains only tokens of type “word” normalized by root (i.e., lemmatized words). Feature selection has not been used [35]. The result is a corporation-specific reputational sentiment (positive / negative) on document level.

## V. EVALUATION

The evaluation compares the classifier’s results to the gold standard, provided by a dataset of reputation-related texts, which have been annotated by humans for reputational sentiment.

The dataset consists of 688 text documents from online media related to corporations’ reputation. The documents were annotated by reputation experts from the banking sector, considering all factors that can influence reputation. Each document was annotated with a fuzzy sentiment label, i.e. each document was annotated with a specific degree of membership to the classes of positive and negative sentiment. The positive and negative membership degrees have five values each with an ascending degree of membership. In this work, binary annotations were derived from fuzzy sentiment labels by the following rule: documents with a higher positive than negative degree of membership is part of the positive class and all other document are part of the negative class. The positive class contains 40% of the documents of our dataset and the negative class 60%.

The dataset was annotated in three rounds: The first round consisted of 269 documents and was annotated by three experts. The dataset was randomly divided among the annotators so that each document was annotated by at least one annotator. In the second round, 394 documents were annotated by four annotators. Again, each document was classified by at least one randomly chosen annotator. In a third round, one annotator annotated 25 documents.

TABLE I.  
CLASSIFIER PERFORMANCE

	Precision	Recall	F-Measure	Accuracy
Positive	87.8%	70.8%	78.4%	84.4%
Negative	84.2%	91.1%	87.5%	84.4%
Micro Avg.	85.4%	83.0%	84.2%	

To evaluate the agreement among annotators for the reputational sentiment annotations, Fleiss’ Kapa inter-rater agreement for nominal scaled values with more than two raters was used [36]. In the first round, 27 documents have been annotated by all three annotators and considering only the positive and negative class, Fleiss’ Kappa of these annotations is 0.78. In the second round, all annotators annotated each of 49 documents. The Fleiss’ kappa from these 49 documents’ annotations is 0.66. We consider the level of agreement fairly well, thus the corpus can be used for evaluation of our classification approach.

Following [38, 37], stratified ten-fold cross validation was used. After classifying every document with the classifier on a test subset, we calculated the standard information retrieval metrics and micro averaged them [39].

Table 1 shows the evaluation results. Our approach could not recognize corporations or sentiment in 19 documents, which were not included in the evaluation. Our accuracy of 84.4% is comparable to results from state of the art sentiment classification research [35], [40].

## VI. DISCUSSION

The contribution of this work is an information retrieval approach for efficiently and comprehensively analyzing corporate reputation automatically from online media texts. Our approach builds upon an ontology for identifying all text parts relating to the same corporation. We contribute a unique dataset of labelled corporation reputation texts (see <https://wi2.uni-hohenheim.de/analytics>) and use it for corporation-specific reputational sentiment analysis by a machine learning method. The evaluation of our approach shows an overall accuracy of 84.4%.

A limitation of this work is that the neutral sentiment orientation is omitted, because [32] have found sentiment classification performance to be substantially higher when omitting the neutral class. We deliberately did not use deep learning techniques because the size of our dataset is too small. That is, the size of our dataset is a limitation. However, human annotation is costly and the size of our dataset is not much smaller than related work (e.g., [32]).

From a managerial perspective, our work helps to efficiently measure corporate reputation in on online world where news and opinions travel fast. Thus, managers can make better decisions by constantly monitoring reputation.

Future work points to comparing our measure of corporate reputation for online media with existing survey-based measures to gain insights about measurement validity. Furthermore, our measure of corporate reputation should be empirically validated by its ability to sense impacts on the financial and economic prospects of a corporation.

## REFERENCES

- [1] A. Pharoah, "Corporate Reputation: The Boardroom Challenge," *Corp. Gov.*, vol. 3, no. 4, pp. 46–51, 2003. <http://dx.doi.org/10.1108/14720700310497113>
- [2] H. Shefrin and M. Statman, "Making sense of beta, size and book-to-market," *J. Portfolio Manage.*, vol. 21, no. 2, pp. 26–34, 1995. <http://dx.doi.org/10.3905/jpm.1995.409506>
- [3] H. Shefrin, "Do investors expect higher returns from safer stocks than from riskier stocks?," *J. Psychol. Financ. Market.*, vol. 2, no. 4, pp. 37–41, 2001. [http://dx.doi.org/10.1207/S15327760JPFM0204\\_1](http://dx.doi.org/10.1207/S15327760JPFM0204_1)
- [4] D. MacGregor, P. Slovic, D. Dreman, and M. Berry, "Imagery, affect, and financial judgment," *J. Psychol. Financ. Market.*, vol. 1, no. 2, pp. 104–110, 2000. [http://dx.doi.org/10.1207/S15327760JPFM0102\\_2](http://dx.doi.org/10.1207/S15327760JPFM0102_2)
- [5] S. Hammond and J. Slocum, "The impact of prior firm financial performance on subsequent corporate reputation," *J. Bus. Ethics.*, vol. 15, no. 2, pp. 159–165, 1996. <https://doi.org/10.1007/BF00705584>
- [6] M. Sobol and G. Farrelly, "Corporate reputation: A function of relative size or financial performance," *Rev. Bus. Econ. Res.*, vol. 24, no. 1, pp. 45–59, 1988.
- [7] P. Roberts and G. Dowling, "Corporate reputation and sustained superior financial performance," *Strateg. Manage. J.*, vol. 23, no. 12, pp. 1077–1093, 2002. <http://dx.doi.org/10.1002/smj.274>
- [8] J. Bloemer, K. De Ruyter, and P. Peeters, "Investigating drivers of bank loyalty: the complex relationship between image, service quality and satisfaction," *Int. J. Bank. Market.*, vol. 16, no. 7, pp. 276–286, 1998. <https://doi.org/10.1108/02652329810245984>
- [9] G. E. Fryxell and J. Wang, "The Fortune Corporate 'Reputation' Index: Reputation for What?," *J. Manage.*, vol. 20, no. 1, pp. 1–14, 1994. <https://doi.org/10.1177/014920639402000101>
- [10] S. Brown, B. Perry, "Removing the Financial Performance Halo from Fortune's 'Most Admired' Companies," *Acad. Manage. J.*, vol. 37, no. 5, pp. 1347–1359, 1994. <https://doi.org/10.5465/256676>
- [11] V. Kubitschek, "Business discontinuity – a risk too far," *Balance Sheet*, vol. 9, no. 3, pp. 33–38, 2001. <http://doi.org/10.1108/09657960110696032>
- [12] C. J. Fombrun and C. B. M. van Riel, "The Reputational Landscape," *Corporate Reputation Review*, vol. 1, no. 1, pp. 5–13, 1997. <https://doi.org/10.1057/palgrave.crr.1540008>
- [13] M. L. Barnett, J. M. Jermier, and B. Lafferty, "Corporate Reputation: The Definitional Landscape," *Corporate Reputation Review*, vol. 9, no. 1, pp. 26–38, 2006. <http://doi.org/10.1057/palgrave.crr.1550012>
- [14] T. J. Brown, P. A. Dacin, M. G. Pratt, and D. . Whetten, "Identity, Intended Image, Construed Image, and Reputation: An Interdisciplinary Framework and Suggested Terminology," *J. Acad. Market. Sci.*, vol. 34, no. 2, pp. 99–106, 2006. <http://doi.org/10.1177/0092070305284969>
- [15] E. G. Love and M. Kraatz, "Character, Conformity, or the Bottom Line? How and Why Downsizing Affected Corporate Reputation," *Acad. Manage. J.*, vol. 52, no. 2, pp. 314–335, 2009. <http://doi.org/10.5465/AMJ.2009.37308247>
- [16] D. Lange, P. M. Lee, and Y. Dai, "Organizational Reputation: A Review," *J. Manage.*, vol. 37, no. 1, pp. 153–184, 2010. <http://doi.org/10.1177/0149206310390963>
- [17] P. Rhee, M. Haunschild, "The liability of good reputation: A study of product recalls in the US automobile industry," *Organization Science*, vol. 17, no. 1, pp. 101–117, 2006. <https://doi.org/10.1287/orsc.1050.0175>
- [18] D. Basdeo, K. Smith, C. M. Grimm, V. P. Rindova, and P. J. Derfus, "The impact of market actions on firm reputation," *Strateg. Manage. J.*, vol. 27, no. 12, pp. 1205–1219, 2006. <http://doi.org/10.1002/smj.556>
- [19] C. Fombrun and M. Shanley, "What's in a Name? Reputation Building and Corporate Strategy," *Acad. Manage. J.*, vol. 33, no. 2, pp. 233–258, 1990. <http://doi.org/10.2307/256324>
- [20] S. J. Brammer and S. Pavelin, "Corporate Reputation and Social Performance: The Importance of Fit," *Journal of Management Studies*, vol. 43, no. 3, pp. 435–455, 2006. <https://doi.org/10.1111/j.1467-6486.2006.00597.x>
- [21] C. Fombrun, "Corporate Reputation—its Measurement and Management," *Thesis*, vol. 18, no. 4, pp. 23–26, 2001.
- [22] D. Turban, D., Greening, "Corporate Social Performance and Organizational Attractiveness to prospective employees," *Acad. Manage. J.*, vol. 40, no. 3, pp. 658–672, 1997. <https://doi.org/10.5465/257057>
- [23] D. Cable and M. Graham, "The determinants of job seekers' reputation perceptions," *J. Organ. Behav.*, vol. 21, no. 8, pp. 929–947, 2000. [https://doi.org/10.1002/1099-1379\(200012\)21:8<929::AID-JOB63>3.0.CO;2-O](https://doi.org/10.1002/1099-1379(200012)21:8<929::AID-JOB63>3.0.CO;2-O)
- [24] V. Rindova and I. Williamson, "Being good or being known: An empirical examination of the dimensions, antecedents, and consequences of organizational reputation," *Acad. Manage. J.*, vol. 48, no. 6, pp. 1033–1049, 2005. <https://doi.org/10.5465/amj.2005.19573108>
- [25] G. Davies, R. Chun, and R. da Silva, "The personification metaphor as a measurement approach for corporate reputation," *Corporate Reputation Review*, vol. 4, no. 2, pp. 113–127, 2001. <https://doi.org/10.1057/palgrave.crr.1540137>
- [26] B. Liu and Zhang, "A survey of opinion mining and sentiment analysis," in *Mining Text Data*, 2012, pp. 415–463. [https://doi.org/10.1007/978-1-4614-3223-4\\_13](https://doi.org/10.1007/978-1-4614-3223-4_13)
- [27] A. Klein, O. Altuntas, T. Haeusser, and W. Kessler, "Extracting Investor Sentiment from Weblog Texts: A Knowledge-based Approach," in 13th Conference on Commerce and Enterprise Computing IEEE, 2011, pp. 1–9. <https://doi.org/10.1109/CEC.2011.10>
- [28] A. Klein, O. Altuntas, M. Riekert, and V. Dinev, "A Combined Approach for Extracting Financial Instrument-Specific Investor Sentiment from Weblogs," in 11th International Conference on Wirtschaftsinformatik, 2013, pp. 691–705.
- [29] F. Sebastiani, "Machine learning in automated text categorization," *ACM Computing Surveys*, vol. 34, no. 1, pp. 1–47, Mar. 2002. <https://doi.org/10.1145/505282.505283>
- [30] T. Joachims, "Text categorization with support vector machines: Learning with many relevant features," in 10th European Conference on Machine Learning, 1998, vol. 1398, no. 2, pp. 137–142. <https://doi.org/10.1007/BFb0026683>
- [31] D. Maynard et al., "Architectural elements of language engineering robustness," *Natural Language Engineering*, vol. 8, pp. 257–274, 2002. <https://doi.org/10.1017/S1351324902002930>
- [32] N. O'Hare et al., "Topic-Dependent Sentiment Analysis of Financial Blogs," in International CIKM Workshop on Topic-Sentiment Analysis for Mass Opinion Measurement, 2009, pp. 9–16. <https://doi.org/10.1145/1651461.1651464>
- [33] B. Pang, L. Lee, and S. Vaithyanathan, "Thumbs up?: sentiment classification using machine learning techniques," in Conference on Empirical Methods in Natural Language Processing, 2002, pp. 79–86. <https://doi.org/10.3115/1118693.1118704>
- [34] M. Riekert, J. Leukel, and A. Klein, "Online Media Sentiment: Understanding Machine Learning-Based Classifiers," *Proceedings of the 24th European Conference on Information Systems (ECIS)*, 2016.
- [35] H. Tang, S. Tan, and X. Cheng, "A survey on sentiment detection of reviews," *Expert Systems with Applications*, vol. 36, no. 7, pp. 10760–10773, Sep. 2009. <https://doi.org/10.1016/j.eswa.2009.02.063>
- [36] J. L. Fleiss, "Measuring nominal scale agreement among many raters," *Psychological bulletin*, vol. 76, no. 5, 1971. <http://doi.org/10.1037/h0031619>
- [37] B. Efron, "Estimating the error rate of a prediction rule: improvement on cross-validation," *Journal of the American Statistical Association*, vol. 78, no. 382, pp. 316–331, 1983. <https://doi.org/10.2307/2288636>
- [38] R. Kohavi, "A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection," in *Proceedings of the 14th International Joint Conference on Artificial Intelligence*, 1995, pp. 1137–1143.
- [39] Y. Yang, "An evaluation of statistical approaches to text categorization," *Information retrieval*, vol. 1, no. 1–2, pp. 69–90, 1999. <https://doi.org/10.1023/A:1009982220290>
- [40] R. Moraes, J. F. Valiati, and W. P. Gavião Neto, "Document-level sentiment classification: An empirical comparison between SVM and ANN," *Expert Systems with Applications*, vol. 40, no. 2, pp. 621–633, Feb. 2013. <https://doi.org/10.1016/j.eswa.2012.07.059>