

Computer-Aided Text Analysis of Corporate Disclosures - Demonstration and Evaluation of Two Approaches

Benjamin Matthies. South Westphalia University of Applied Sciences. Germany

matthies.benjamin@fh-swf.de

André Coners. South Westphalia University of Applied Sciences. Germany

Abstract. The volume of corporate disclosure is constantly growing and increasing attention is paid to the systematic exploration of its highly informative textual content. Manual analyses, however, are quickly reaching their capacity limits when exploring large collections of texts. Computer-aided text analyses are therefore becoming increasingly important in order to overcome the information overload. In accounting research, however, the corresponding possibilities and limitations of such computer-based analyses are hardly discussed. This paper addresses this knowledge gap and pursues the goal of demonstrating the use of computer-aided text analysis approaches and providing concrete recommendations of “dos” and “don'ts” for their application. Within the framework of a case study, two text analysis strategies – dictionary and statistical approach – are practically applied, documented and subsequently discussed. In conclusion, computer-based processes have proven to be an efficient means for coping with large text collections. Furthermore, the combined use of both text analysis approaches has proven advantageous since they complement each other and compensate for each other's weaknesses. The combination of quantitative results related to thematic categories (dictionary approach) as well as the exploration of new content patterns (statistical approach) provides a more comprehensive picture with regard to the presentation of corporate disclosure.

Keywords: corporate disclosure, computer-aided text analysis, text mining.

1. INTRODUCTION

The volume of corporate disclosure is constantly growing (OECD, 2012) and contains much more than just financial statements nowadays. To an increasing extent, companies publish additional descriptions of their business strategies, their objectives, risks and opportunities as well as their intangible resources and abilities (Bowman, 1984; Dirsmith & Covalski, 1983; Healy & Palepu, 2001; Schlüter *et al.*, 2014; WGARIA, 2005). Take Volkswagen AG as an example. The annual report in 2003 was 188 pages long, but grew to 494 pages in 2013 – whereby only 116 pages cover the financial statements (i.a., balance sheet, income statement), while 161 pages contain the discussion of business management issues (i.a., strategy, business development, risks and opportunities). This illustrates the information potential of such disclosures and, consequently, more and more attention is paid to the systematic exploration of this usually textual content (Beattie & Thompson, 2007; Bowman, 1984; Guthrie *et al.*, 2004; Healy & Palepu, 2001; Li, 2010). The goals of such explorations are to determine the reported topics, their volume and dissemination as well as the trends in the discussed issues (Beattie & Thompson, 2007; Li 2010). The extraction of patterns can provide important evidences about the current fields of action in corporate practice and also offer insight into management's decisions and incentives (Li, 2010).

The growing volume of corporate disclosure, however, involves not only great information potential, but also great challenges. Research and practice are increasingly facing the problem of finding relevant patterns in the large amounts of textual reports, structuring them and extracting useful knowledge – of course, all this in a reasonable amount of time. Manual approaches are therefore quickly reaching their capacity limits when analyzing large collections of texts (Indulska *et al.*, 2012). Consequently, computer-aided text analyses are becoming increasingly important (Li, 2010; Morris, 1994) and could help to overcome the “crisis of information overload” (Feldman & Sanger, 2007). In accounting research, however, the corresponding possibilities and limitations or likewise the advantages and disadvantages of such computer-aided analyses are hardly discussed and comprehensively summarized, which is why their potential is still difficult to assess. This paper addresses this knowledge gap and considers the following research question:

How can textual corporate disclosures be examined with computer-aided text analyses and what is necessary to consider in their use?

To answer this question, it is first of all necessary to look at the theoretical background of corporate disclosure and the potentials for its textual analysis (Section 2). Then two approaches for a computer-aided text analysis (dictionary and statistical approach) will be introduced (Section 3). This will be followed by an exploratory case study to demonstrate their application (Section 4). In this demonstrative analysis, the risk reporting in the German industrial sector is examined before, during and after the global financial crisis (2007-2008). 114 annual reports with a total of more than 15,000 pages will be analyzed mostly automatically. In this process, the focus is mainly on demonstrating and documenting the use of the introduced computer-aided text analysis approaches. The gained experiences shall be discussed, evaluated and concrete recommendations for using such computer-assisted means shall then be made (Section 5). In conclusion, the possibilities and limitations of computerized text analyses will be summarized for the accounting discipline (Section 6).

2. TEXTUAL ANALYSIS OF CORPORATE DISCLOSURES

The external recipients of corporate disclosures, such as financial analysts and potential investors, are especially interested in the future potential of companies. This, however, is only insufficiently forecasted by the financial performance, since such statements are primarily based on historic transactions (Alwert *et al.*, 2008; FEI, 2001; Healy & Palepu, 2001; Hooks *et al.*, 2002). Multiple institutions in accounting have taken up this growing need for information and are concretely demanding to include additional descriptions of the corporate strategy, business development, future risks and opportunities as well as intangible resources and abilities in the reports (AICPA, 1994; APCA, 2010; FASB, 2001; FEI, 2001; IASB, 2010). The voluntary and usually textual description of such issues complements the traditional “financial reporting” by a so-called “business reporting” (AICPA, 1994; FASB, 2001). By this means, financial analysts and potential investors should be able to make a more transparent assessment of companies, which will reduce the information asymmetry, help with investment decisions and consequently contribute to a more realistic convergence between the market value and the actual corporate value (Healy & Palepu, 2001). The

publishing companies of such a value reporting profit from the fact that they could receive capital more efficiently, reduce their transaction and capital costs, and ultimately can increase their market value (Boesso & Kumar, 2007; Botosan, 1997; Francis *et al.*, 2008; Healy & Palepu, 2001). The analysis of textual corporate disclosures has been widespread for a long time now and offers insight into the strategies of companies (Bowman, 1984; Dirsmith & Covalesski, 1983), their risks and opportunities (Schlüter *et al.*, 2014), their intangible resources and abilities (Beattie & Thompson, 2007; Guthrie *et al.*, 2004) as well as their corporate social responsibility efforts (Gamerschlag *et al.*, 2011). In this regard, such analyses can be conducted for several purposes:

- The analysis of the disclosed reports, their volume, and their linguistic structure can supply research with a rich database for the empirical examination of known theories in corporate reporting (e.g., signaling theory, agency theory, or legitimacy theory; see, e.g., Campbell *et al.*, 2001).
- The analysis of textual disclosures represents a supplementary function for the interpretation of financial data. It reveals how financial performance indicators are composed and helps with understanding the financial background and prospects (Beynon *et al.*, 2004; Li, 2010).
- The analysis of the publications by competitors offers insight into their strategies (e.g., market entries) and intended actions (e.g., organizational restructurings) as well as their business success and failure. By this means, such analyses can provide valuable information in the sense of a competitor accounting (Subramanian & IsHak, 1998).
- The analysis of textual disclosures sheds light on managers' decisions and reasoning, improving the understanding of management characteristics and incentives (Li, 2010).
- The analysis of textual corporate disclosures can provide risk-related predictors with respect to, for example, market failure (Sarlin & Marghescu, 2011), corporate bankruptcy (Shirata *et al.*, 2011), or fraud (Goel & Gangolly, 2012).

3. COMPUTER-AIDED TEXT ANALYSIS APPROACHES

The manual analysis of textual disclosures has been widely applied for a long time now (Beattie & Thompson, 2007; Healy & Palepu, 2001). However, manual approaches are quickly reaching their capacity limits when analyzing large collections of texts (Indulska *et al.*, 2012). For this reason, the computer-aided analysis of texts is increasingly in focus (Morris, 1994; Li, 2010). The technical foundations of such analyses have made great progress in the last two decades, driven by constant innovation in complementary fields such as computational linguistics, databases, information retrieval, or data mining (Miner *et al.*, 2012). A clear generic definition for computerized text analyses has not yet established itself and various names are still frequently used for this, such as “text mining”, “text analytics” or “knowledge discovery in textual databases” (Gao *et al.*, 2005). Fundamentally, however, all these names share the common approach of “turning texts into numbers” (Miner *et al.*, 2012), i.e. breaking down unstructured textual reports into individual words and word combinations, counting them, assigning them to the respective documents, and finally transferring them to a structured numeric format (Manning & Schütze, 1999). On the basis of this data structure, it becomes possible to record and organize extensive collections of texts in a mostly automatic way, to retrieve specific content (e.g., words, word combinations, or sentences), to investigate unknown structures in documents and document groups, to discover identical or correlated content, and thus to generate relevant knowledge (Feldman & Sanger, 2007; Miner *et al.*, 2012; Weiss *et al.*, 2010). Besides their efficiency, other advantages of computerized analyses are that they can be replicated at any time, are always reliable under the given settings and lead to the same results. This means, in contrast to manual processes, that are not dependent on the subjective interpretation or attention of the analyst (Indulska *et al.*, 2012).

Fundamentally, it is possible to differentiate between two text analysis approaches in this context: a rule-based dictionary approach and an exploratory statistical approach (Li, 2010). Since the demonstration and evaluation of these two approaches is at the center of this paper, they shall be introduced in the following:

Dictionary approach: Just as surveys or interviews can be structured for the generation of knowledge, the analysis of textual content can also be structured. A

popular approach is the development of so-called “dictionaries”, which thematically model the text analysis on the basis of predefined thematic categories and keywords (Li, 2010). The keywords are counted in the underlying documents, and their frequency is considered a general indicator for their relevance or, in the case of a summary, for the relevance of the respective thematic categories (Krippendorff, 2013; Weber, 1990). Therefore, the dictionary approach is a conceptual analysis that evaluates texts on the basis of the existence and frequency of specific content and thereby produces statements with regard to the volume, dissemination, and development of predefined thematic concepts (Indulska *et al.*, 2012).

Statistical approach: In some cases, attention should be paid not only to what is disclosed but also to how it is disclosed. For this purpose, statistical analyses are well-suited for investigating contextual characteristics of the discussion. Statistical text analyses are primarily based on exploratory processes that aim to extract relevant and previously unknown knowledge from textual databases (Li, 2010). Such statistical approaches analyze texts mathematically, without predefined limitations, and can discover connections in the content by revealing statistical structures or patterns. Accordingly, it is possible to classify or cluster various documents by their match in terms of content. Furthermore, correlations between keywords which are often connected with each other can be discovered (e.g., by joint appearance in sentences) and thematic concepts can be illustrated on the basis of visualization techniques (e.g., association matrices). Therefore, the statistical approach is essentially a relational analysis that not only calculates the frequency of content, but also discovers the co-occurrence of concepts, i.e. content connections and relationship structures in texts (Indulska *et al.*, 2012).

As was mentioned at the outset, computer-aided text analysis is increasingly in focus and some studies already provide general insights into its use in the context of accounting research. A few specific examples should be given. Gamerschlag *et al.* (2011) used the dictionary approach (two thematic categories with a total number of 32 keywords) in order to extract German companies' Corporate Social Responsibility (CSR) disclosures in corporate reports. Shirata *et al.* (2011) performed a statistical analysis of financial reports in Japan and extracted key phrases to predict bankruptcy. Kloptchenko *et al.* (2004) used the statistical

approach in order to explore the linguistic style and contents of quarterly financial reports. Goel and Gangolly (2012) performed statistical analyses of textual content in annual reports in order to explore differences in the writing and presentation style between companies that committed fraud and those that did not. Nevertheless, to date, the corresponding possibilities and limitations of such computer-assisted processes are hardly discussed in the accounting literature and have not been comprehensively summarized. This paper addresses this knowledge gap in the following case study and produces such an overall picture.

4. CASE STUDY

4.1 Case Study Design

An exploratory case study approach is chosen to demonstrate and evaluate the application of computerized methods for analyzing textual corporate disclosures. In this context, two approaches to computer-aided text analysis – dictionary and statistical approach – are practically used, documented and subsequently discussed. Such an “application description” represents a special form of case study research, the goal of which is to demonstrate the application or implementation of certain information technology processes or systems and to make concrete recommendations on the basis of the gained experiences (Benbasat *et al.*, 1987). The authors of this paper conducted the case study themselves and describe their own experiences, which is why this procedure can be categorized as action research (McNiff & Whitehead, 2000).

The case study shall be presented in more detail below. It relies on the steps of a computer-aided text analysis process (see Figure 1) proposed by Miner *et al.* (2012, p. 75).

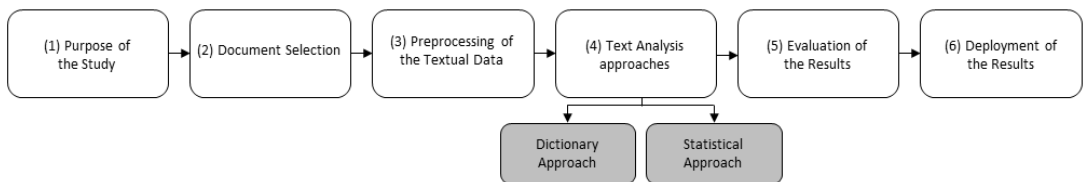


Figure 1. Steps of a computer-aided text analysis process.

(1) Purpose of the Study: Risk reporting has attracted greater interest in the past few years. In particular, this is due to a stronger risk awareness after the global financial crisis (Schlüter *et al.*, 2014) as well as the fact that multiple regulations require more comprehensive disclosure of corporate risks and opportunities (e.g., GAS 20; IFRS 7; or FASB ASC 275). This paper takes up this subject to exemplify computer-aided text analyses and examines in particular the risk reporting before, during and after the global financial crisis (2007-2008). There are already some studies available that pursued a comparable goal (Probohudono *et al.*, 2013; Schlüter *et al.*, 2014). However, the study of this paper relies completely on automated and exploratory text analysis techniques. Furthermore, the focus of this study is on the Germany-listed DAX companies in the industrial sector (according to the definition of the Industry Classification Benchmark; see FTSE, 2013). This sector is interesting because it plays a central role for German economy and had tremendous difficulties during the global crisis due to declines in orders and financial difficulties. This is why the concrete goal of this study is to examine the risk awareness and publicity of the German industrial sector before, during and after the crisis. Or in other words: How did the financial/economic crisis impact the risk reporting of the German industrial sector? This general question is the starting point for examining three more concrete research questions in the following analyses:

RQ1: Which types of risk took precedence in the risk discussion?

RQ2: Did the volume of risk reporting increase during and after the crisis?

RQ3: In what context were companies' risks reported on?

It is important to note that the focus of this study is mainly on demonstrating and evaluating the use of computerized text analyses and rather less on conducting a full empirical study. Therefore, a more detailed study design (with hypotheses development and verification) has not been considered as necessary.

(2) Document Selection: The annual reports of publicly-listed DAX companies in the industrial sector served as the database. The reporting periods of 2006 (beginning of the financial crisis), 2009 (peak of the following economic crisis in the industrial sector) and 2012 (after the crisis) were exclusively taken into account. A comparison of these reporting periods should show the changes

and trends in the risk reporting. Only English reports were included and downloaded from the companies' websites. Since some annual reports were not available online or were not present in the required form, there was ultimately a database of 114 annual reports (38 per year) with more than 15,000 pages.

Annual reports are especially well-suited for the analytical purpose of this study, since they represent the most important and most complete instrument for companies to communicate with external addressees (Dirsmith & Covaleski, 1983; Healy & Palepu, 2001). Furthermore, they are fundamentally subject to the disclosure requirements with regard to the discussion of risks (e.g., GAS 20; IFRS 7), which ensures their widespread availability and relatively good comparability.

(3) Preprocessing of the Textual Data: In preparing the actual data, various automated processes were performed, such as: word extraction, stop-word cleaning, and lemmatization (Manning & Schütze, 1999). These processes break down the unstructured text content into individual words and word combinations, count them, assign them to individual documents and thereby ultimately transfer them to an analyzable data structure. The resulting term-document matrix (TDM) forms the structured, i.e. relational and numeric database for the following text analysis approaches (see an example in Figure 2). In this process, only the sections that contain discussions about risks were selected (between 2 and 18 pages per report). Other irrelevant content in the annual reports (e.g., company presentations) are excluded in the automatic data preparation in order to avoid distortions in the following statistical analyses. Finally, a manual evaluation and adjustment of this database had to be done (e.g., elimination of irrelevant terms and correction of ambiguous words, misspellings or obviously wrong processed data). The effort for these data corrections should not be underestimated.

(4) Textual Analysis Approaches: After the texts were transferred to a TDM, the created database could be analyzed on the basis of various processes.¹ In the context of this study, the two already mentioned approaches – dictionary and statistical approach – were applied. The execution and the results of these two approaches are described in the following Sections 4.2 and 4.3.

(5) Evaluation of the Results: The evaluation of the results refers primarily

¹ This study used *Provalis Research's* QDA Miner 4 with its text mining-application WordStat 7.

to the documentation and discussion of the applied text analysis processes. In Section 5, the experiences will be presented and the respective the possibilities and limitations or likewise the advantages and disadvantages of the approaches will be discussed.

(6) Use of the Results: The findings of the case study are used in Section 6 to provide conclusive recommendations for accounting research and practice.

	Amadeus_2006	Amadeus_2009	Amadeus_2012	Balda_2006	Balda_2009	Balda_2012	Baywa_2006	Baywa_2009	Baywa_2012	BDI_2006	BDI_2009	BDI_2012	Bill
CAPITAL			1	7	8	5	2	5	3				
COMPLIANCE	1	4	5			3		5	4				
CONTROL	3	8	11	6	13	40	2	16	13				
COST				5	4	4		1	3	1			2
CREDIT					3	1	2	3	5	1	1		1
CURRENCY	1	1		2	4	2	5	2	2		2		1
CUSTOMER			5	22	26	23	1	2	2	2	4		
DEVELOPMENT	6	7	5	18	23	20	11	9	12		4		4
EARNING	1	2	2	4	1	1		1			1		1
ECONOMIC	7	5	7	15	19	7	4	2	4	2	1		1
ECONOMY	2	5	6	5	12	3	2	1			1		1
EMPLOYEE	3	7	12	2	2	2	2	12	9		2		1
FINANCE	3	4	4	8	3	3	9	8	13				1
FINANCIAL	4	21	30	15	14	29	3	22	12		6		5
GLOBAL	1		2	13	20	7	3	3	3	1	3		4
GROWTH	3	1	5	17	11	5	4			3	1		1
HEDGE				1	1	2	3	2	4				1
INTEREST				1	6	1	1	5	3				
INTERNATIONAL				7	10	12	1	1	2	2	3		1
INVESTMENT		1	1	2		6	4	5	5				
LIMIT	1	1	1	2	3	3	3	5	4				
MARKET	6	4	6	26	33	24	10	8	9	1	8		1
MATERIAL	1	1	2	4	6	3	1	5			9		4
OPERATION	1	3	4	4			5	2	3	1	5		3
PERFORMANCE		2	4	1	7	2	2	4	4		1		1
POLICY	3	2	5	1	3	2	2	2	6		1		1
PRICE	2			12	14	5	4	11	10		1		
PRODUCT				10	14	10	1	2	2				

Figure 2. Term-document matrix (TDM).

4.2 Dictionary Approach

4.2.1 Design

In the dictionary approach, the object of investigation must be modelled thematically on the basis of meaningful categories and keywords. For this study, therefore, appropriate categories must be created in the first step, structuring the risk reporting and representing it as complete, objective and valid as possible. In order to ensure the objectivity and validity of these categories, Krippendorff (2013) recommended relying on the so-called “sources of certainty” such as expert knowledge, substantiated theories, and established practical models. Therefore, this study relied on the comprehensive risk categories of the Enterprise Risk Management Committee hosted by the *Casualty Actuarial Society* (CAS,

2003, p. 10): hazard risks, financial risks, operational risks and strategic risks. In order to record financial risks in a more granular and informative way, they were further specified as market risks (market price risks), credit risks and liquidity risks (referring to GAS 20). In the next step, it was necessary to define suitable keywords and to assign them to the thematic categories. Here it is also recommended to consider the most reliable references (Krippendorff, 2013). For the identification of valid keywords, therefore, a review of the relevant regulations and risk management concepts was conducted (CAS, 2003; GAS 20; IFRS 7; Global Report Initiative; Art. 315 of the German Commercial Code [HGB]; FASB Accounting Standards Codification Topic 275). The complete dictionary with all the categories and their respective keywords is provided in Table 1.

The modelling and analytical implementation of such dictionaries is supported by the common text analysis tools. The categories and sub-categories are initially organized with a folder structure until individual keywords and word combinations can be assigned. The conventional tools offer the analyst various forms of support for specifying the keyword search. In order to take the diverse orthography and means of expression into account, searchable terms can be reduced to their word stem (e.g., “disclos*” for “disclosure” or “disclosing”). The analyst is also offered synonyms automatically for an expansion of the keywords (e.g., “liability” and “obligation” for “debt”). In addition, the definition of individual search rules is possible in order to specify the word search (e.g., AND, OR, WITHOUT, NOT AFTER/BEFORE). If need be, a different weighting of individual keywords can be implemented (e.g., for the calculation of indices). In order to ensure the suitability of the identified keywords, tests such as keyword-in-context (KWIC) are offered. Keywords are systematically retrieved in their respective sentences (context) and displayed, which makes an evaluation of their actual meaning easier. The subsequent counting of the keywords takes place automatically, and the frequencies are summarized in accordance with the respective categories. The results are available to the analyst in multiple forms (e.g., absolute frequencies or percentage distribution). Furthermore, various statistical evaluations (e.g., regression analysis) and tests (e.g., ANOVA) are also usually available in the common tools.

Categories	Keywords
Hazard risks	hazard, damage, natural perils, windstorms, theft, crime, injury, disease, disability, chemical hazards, biological hazards, toxic, workplace accidents, waste, pollution, fraud, illegal
Financial risks	
Market risks	market risk, price change, interest rate, equity risk, currency, foreign exchange rate, commodity price, option risk, volatility, property risk, inflation, deflation
Credit risks	credit risk, fair value change, concentration risk, contingency risk, default risk, credit rating risk, credit spread, sovereign risk, transfer risk, hedging, downgrade, liability, loans payable, loan default, solvency
Liquidity risks	liquidity risk, call risk, deadline risk, market liquidity risk, payment flow, payment obligations
Operational risks	operational risk, empowerment, information technology, data security, capacity, personnel risk, staff risk, system risk, systemic risk, litigation risk, internal risk, transaction exposure, position risk, business risk, labor turnover, business interruption, product development, project risks, product/service failure, human resources, fluctuation
Strategic risks	strategic risk, corporate/ business strategy, reputational damage, competition, customer wants, warranty, demographic, social/cultural trends, technological innovations, regulatory/political, environmental risk, compensation claims, policy, economy, R&D, patents, legal risk, government

Notes:

The risk categories have been derived from the Enterprise Risks Management (ERM) Framework (CAS, 2003).

The keywords derived from a review of CAS; GAS 20; Art. 315 HGB; IFRS 7; Global Report Initiative; FASB ASC 275

Table 1. Dictionary for exploring risk disclosures

4.2.2 Results

In the following, the quantitative results of the dictionary approach are presented. Table 2 provides the descriptive statistics with respect to the volume and dissemination of the risk-related disclosures by risk category. A trend analysis is summarized in Table 3.

Categories	Total	% of reports	Mean	S.D.	Min.	Max	Median
Hazard risks	449	77%	5.10	4.58	1	20	4
Financial risks							
Market risks	2,333	99%	20.65	14.50	1	106	18
Credit risks	2,350	100%	20.61	14.63	1	83	18
Liquidity risks	1,002	95%	9.28	7.03	1	34	8
Operational risks	1,153	97%	10.39	6.50	1	29	9
Strategic risks	3,385	98%	30.22	19.88	3	125	27
Total	10,672	100%	93.61	53.95	4	289	85

Notes: S.D.: standard deviation; Min: minimum of disclosure amount; Max: maximum of disclosure amount per report

Table 2. Descriptive statistics of risk disclosure per category

Overall, the automatic counting process identified 10,672 keywords in the companies' risk reporting (mean: 93.61). An in-depth analysis suggests that the discussion of the examined risk categories is available in almost all reports (> 95% of reports). Only hazard risks are mentioned less, in 77% of the reports. In terms of the mean values² of the risk categories, it becomes clear that financial risks take up the largest amount in the risk reports (means: market risks = 20.65; credit risks = 20.61; liquidity risks = 9.28), followed by strategic risks (mean: 30.22) and operational risks (mean: 10.39). Hazard risks are mentioned significantly less in the reports (mean: 5.10). In this connection, however, reference must be made to the partially high standard deviations. With regard to the first research question in the analysis (Which types of risk took precedence in the risk discussion?), the quantitative results of the dictionary approach help to confirm that financial and strategic aspects take precedence in the risk discussion, followed by operational aspects and hazards.

² A varying number of report pages has not been considered since the overall amount of disclosures is in focus of this study rather than an average value per report page.

Categories	2006	2009	2012
Hazard risks	3.26	5.25***	6.24
Financial risks			
Market risks	17.95	20,87***	23.19
Credit risks	13.79	22.50***	25.55*
Liquidity risks	5.53	10.41***	11.59
Operational risks	7.17	11.11***	12.74*
Strategic risks	17.89	30.87***	41.26***
Total	62.39	99.61***	118.84**

Notes: ***/**/* = 1%-, 5%-, 10%-level of significance of increase compared with previous period (non-parametric Wilcoxon signed-rank test for two related samples; a normal distribution assumption cannot be completely supported).

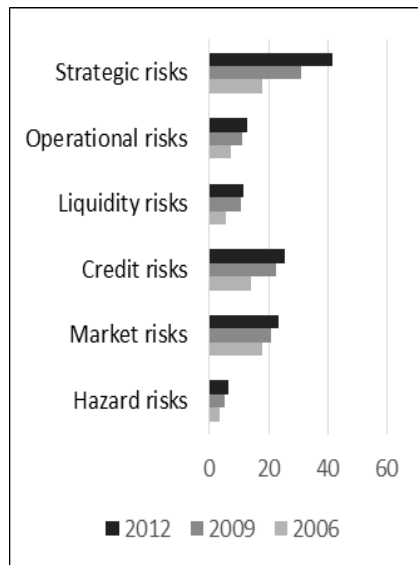


Table 3. Trend analysis of risk disclosures

In general, the trend analysis (see Table 3) confirmed that the volume of reporting on the respective risk categories has increased significantly over the years (tested by a non-parametric Wilcoxon signed-rank test). If one takes into account the change from 2006 (beginning of the financial crisis) to 2009 (peak of the related economic crisis in the industrial sector) and then to 2012 (after the crisis), it becomes clear that the volume of risk reporting increased greatly between 2006 and 2009 in particular (hazard risk + 61%; market risks + 16%; credit risks + 63%; liquidity risks + 88%; operational risks + 55%; strategic risks + 73%). The increase in the volume of reporting is not as great between 2009 and 2012 (hazard risk + 19%; market risks + 11%; credit risks + 14%; liquidity risks + 11%; operational risks + 15%; strategic risks + 34%). Aside from the strategic risks, the increases can also not be viewed as strong significant.

With regard to the second research question (Did the volume of risk reporting increase during and after the crisis?), the quantitative results of the dictionary approach clearly provide information about how risk reporting increased greatly and significantly during and after the crisis. The large reports in subsequent years (2012) were retained, but no noticeable and significant increases can be proven. These trends largely correspond to findings obtained from comparable studies (Schlüter *et al.*, 2014), which confirms their validity to a certain degree.

4.3 Statistical Approach

4.3.1 Design

The statistical approach can uncover previously unknown content relationships and, therefore, make it possible to overcome the fixed framework of the dictionary approach. An entire series of diverse techniques, processes, and visualizations are available as part of such exploratory text analysis (Feldman & Sanger, 2007; Miner *et al.*, 2012; Weiss *et al.*, 2010). In this demonstration, the approach shall be limited to the following four analyses:

Keyword distribution: This analysis provides the most frequent used words and word combinations in the entire document collection. The focus is less on the absolute frequencies of the keywords than on the repeated occurrence of the words (in % of cases) discussed in the various risk reports.

Proximity plots: A proximity plot represents the words that are used most

frequently in connection (in the same sentence or paragraph) with a central term (“risk”) and thus in a strong content relationship. The statistical proximity (cosine statistics; see Feldman & Sanger, 2007, p. 85) of the words that stand next to this term is calculated and shown on a scale of 0-1.

Topic extraction: Topic extraction statistically explores the abstract topics that occur in a document collection and, thereby, uncovers the hidden thematic structure in textual content. A “topic” hereby represents a cluster of words that frequently occur together in connection with a few central terms. The results include the variance of the topics, their distribution within the document collection, and the summarized frequency of the respective keywords.

Link analysis: A link analysis is an instrument for discovering and visualizing relationship patterns in the content of text collections. The combinations of keywords frequently occurring together are statistically analyzed (cosine statistics) and shown in content concepts. Furthermore, a link analysis can explore the statistical associations with a central term and creates corresponding association rules (shown on a scale of 0-1). This technique is comparable to the above mentioned proximity plots, with the difference that interactive network graphs allow to explore content components of the discussions and their contextual connections more visually.

4.3.2 Results

The following results show that the statistical text analysis goes beyond the summarizing statements of the dictionary approach and, with respect to the third research question (In what context were companies' risks reported on?), can explore previously unknown information about the contextual discussion of risks. Such exploratory analyses could therefore have better chances of producing early warnings of beginning crises.

Year	Keyword distribution	Proximity plots																																																																		
2006	<p>2006 Keyword Distribution Data (Approximate):</p> <table border="1"> <tr><th>Keyword</th><th>Items</th></tr> <tr><td>FINANCIAL RISK</td><td>75</td></tr> <tr><td>CURRENCY RISK</td><td>55</td></tr> <tr><td>LEGAL RISK</td><td>52</td></tr> <tr><td>FINANCIAL INSTRUMENT</td><td>50</td></tr> <tr><td>RAW MATERIAL</td><td>48</td></tr> <tr><td>MARKET RISK</td><td>45</td></tr> <tr><td>ECONOMIC GROWTH</td><td>42</td></tr> <tr><td>CAPITAL MARKET</td><td>40</td></tr> <tr><td>GLOBAL ECONOMY</td><td>38</td></tr> <tr><td>ECONOMIC RISK</td><td>35</td></tr> <tr><td>PERSONNEL RISK</td><td>32</td></tr> <tr><td>MATERIAL PRICE</td><td>30</td></tr> <tr><td>PRICE PRESSURE</td><td>28</td></tr> <tr><td>GLOBAL ECONOMIC</td><td>25</td></tr> <tr><td>RESEARCH AND DEVELOPMENT</td><td>22</td></tr> <tr><td>LEGAL DISPUTE</td><td>20</td></tr> </table>	Keyword	Items	FINANCIAL RISK	75	CURRENCY RISK	55	LEGAL RISK	52	FINANCIAL INSTRUMENT	50	RAW MATERIAL	48	MARKET RISK	45	ECONOMIC GROWTH	42	CAPITAL MARKET	40	GLOBAL ECONOMY	38	ECONOMIC RISK	35	PERSONNEL RISK	32	MATERIAL PRICE	30	PRICE PRESSURE	28	GLOBAL ECONOMIC	25	RESEARCH AND DEVELOPMENT	22	LEGAL DISPUTE	20	<p>2006 Proximity Plot Data (Approximate):</p> <table border="1"> <tr><th>Keyword</th><th>RISK</th></tr> <tr><td>FINANCIAL</td><td>0.85</td></tr> <tr><td>MARKET</td><td>0.80</td></tr> <tr><td>DEVELOPMENT</td><td>0.75</td></tr> <tr><td>SALE</td><td>0.72</td></tr> <tr><td>INTEREST</td><td>0.70</td></tr> <tr><td>SERVICE</td><td>0.68</td></tr> <tr><td>PRICE</td><td>0.65</td></tr> <tr><td>CUSTOMER</td><td>0.62</td></tr> <tr><td>HEDGE</td><td>0.60</td></tr> <tr><td>CURRENCY</td><td>0.58</td></tr> <tr><td>CONTROL</td><td>0.55</td></tr> <tr><td>COST</td><td>0.52</td></tr> <tr><td>GROWTH</td><td>0.50</td></tr> <tr><td>EXCHANGE</td><td>0.48</td></tr> <tr><td>LIMIT</td><td>0.45</td></tr> </table>	Keyword	RISK	FINANCIAL	0.85	MARKET	0.80	DEVELOPMENT	0.75	SALE	0.72	INTEREST	0.70	SERVICE	0.68	PRICE	0.65	CUSTOMER	0.62	HEDGE	0.60	CURRENCY	0.58	CONTROL	0.55	COST	0.52	GROWTH	0.50	EXCHANGE	0.48	LIMIT	0.45
Keyword	Items																																																																			
FINANCIAL RISK	75																																																																			
CURRENCY RISK	55																																																																			
LEGAL RISK	52																																																																			
FINANCIAL INSTRUMENT	50																																																																			
RAW MATERIAL	48																																																																			
MARKET RISK	45																																																																			
ECONOMIC GROWTH	42																																																																			
CAPITAL MARKET	40																																																																			
GLOBAL ECONOMY	38																																																																			
ECONOMIC RISK	35																																																																			
PERSONNEL RISK	32																																																																			
MATERIAL PRICE	30																																																																			
PRICE PRESSURE	28																																																																			
GLOBAL ECONOMIC	25																																																																			
RESEARCH AND DEVELOPMENT	22																																																																			
LEGAL DISPUTE	20																																																																			
Keyword	RISK																																																																			
FINANCIAL	0.85																																																																			
MARKET	0.80																																																																			
DEVELOPMENT	0.75																																																																			
SALE	0.72																																																																			
INTEREST	0.70																																																																			
SERVICE	0.68																																																																			
PRICE	0.65																																																																			
CUSTOMER	0.62																																																																			
HEDGE	0.60																																																																			
CURRENCY	0.58																																																																			
CONTROL	0.55																																																																			
COST	0.52																																																																			
GROWTH	0.50																																																																			
EXCHANGE	0.48																																																																			
LIMIT	0.45																																																																			
2009	<p>2009 Keyword Distribution Data (Approximate):</p> <table border="1"> <tr><th>Keyword</th><th>Items</th></tr> <tr><td>FINANCIAL RISK</td><td>85</td></tr> <tr><td>FINANCIAL INSTRUMENT</td><td>75</td></tr> <tr><td>CURRENCY RISK</td><td>70</td></tr> <tr><td>ECONOMIC CRISIS</td><td>65</td></tr> <tr><td>LEGAL RISK</td><td>60</td></tr> <tr><td>GLOBAL ECONOMIC</td><td>55</td></tr> <tr><td>RAW MATERIAL</td><td>50</td></tr> <tr><td>FINANCIAL CRISIS</td><td>48</td></tr> <tr><td>CREDIT RISK</td><td>45</td></tr> <tr><td>CAPITAL MARKET</td><td>42</td></tr> <tr><td>GLOBAL ECONOMY</td><td>40</td></tr> <tr><td>LIQUIDITY RISK</td><td>38</td></tr> <tr><td>MARKET RISK</td><td>35</td></tr> <tr><td>HUMAN RESOURCE</td><td>32</td></tr> <tr><td>ECONOMIC DOWNTURN</td><td>30</td></tr> </table>	Keyword	Items	FINANCIAL RISK	85	FINANCIAL INSTRUMENT	75	CURRENCY RISK	70	ECONOMIC CRISIS	65	LEGAL RISK	60	GLOBAL ECONOMIC	55	RAW MATERIAL	50	FINANCIAL CRISIS	48	CREDIT RISK	45	CAPITAL MARKET	42	GLOBAL ECONOMY	40	LIQUIDITY RISK	38	MARKET RISK	35	HUMAN RESOURCE	32	ECONOMIC DOWNTURN	30	<p>2009 Proximity Plot Data (Approximate):</p> <table border="1"> <tr><th>Keyword</th><th>RISK</th></tr> <tr><td>FINANCIAL</td><td>0.90</td></tr> <tr><td>MARKET</td><td>0.90</td></tr> <tr><td>PRICE</td><td>0.88</td></tr> <tr><td>CUSTOMER</td><td>0.85</td></tr> <tr><td>CONTROL</td><td>0.82</td></tr> <tr><td>DEVELOPMENT</td><td>0.80</td></tr> <tr><td>COST</td><td>0.78</td></tr> <tr><td>ECONOMIC</td><td>0.75</td></tr> <tr><td>SERVICE</td><td>0.72</td></tr> <tr><td>CAPITAL</td><td>0.70</td></tr> <tr><td>EMPLOYEE</td><td>0.68</td></tr> <tr><td>INVESTMENT</td><td>0.65</td></tr> <tr><td>LIABILITY</td><td>0.62</td></tr> <tr><td>TRADE</td><td>0.60</td></tr> <tr><td>POLICY</td><td>0.58</td></tr> </table>	Keyword	RISK	FINANCIAL	0.90	MARKET	0.90	PRICE	0.88	CUSTOMER	0.85	CONTROL	0.82	DEVELOPMENT	0.80	COST	0.78	ECONOMIC	0.75	SERVICE	0.72	CAPITAL	0.70	EMPLOYEE	0.68	INVESTMENT	0.65	LIABILITY	0.62	TRADE	0.60	POLICY	0.58		
Keyword	Items																																																																			
FINANCIAL RISK	85																																																																			
FINANCIAL INSTRUMENT	75																																																																			
CURRENCY RISK	70																																																																			
ECONOMIC CRISIS	65																																																																			
LEGAL RISK	60																																																																			
GLOBAL ECONOMIC	55																																																																			
RAW MATERIAL	50																																																																			
FINANCIAL CRISIS	48																																																																			
CREDIT RISK	45																																																																			
CAPITAL MARKET	42																																																																			
GLOBAL ECONOMY	40																																																																			
LIQUIDITY RISK	38																																																																			
MARKET RISK	35																																																																			
HUMAN RESOURCE	32																																																																			
ECONOMIC DOWNTURN	30																																																																			
Keyword	RISK																																																																			
FINANCIAL	0.90																																																																			
MARKET	0.90																																																																			
PRICE	0.88																																																																			
CUSTOMER	0.85																																																																			
CONTROL	0.82																																																																			
DEVELOPMENT	0.80																																																																			
COST	0.78																																																																			
ECONOMIC	0.75																																																																			
SERVICE	0.72																																																																			
CAPITAL	0.70																																																																			
EMPLOYEE	0.68																																																																			
INVESTMENT	0.65																																																																			
LIABILITY	0.62																																																																			
TRADE	0.60																																																																			
POLICY	0.58																																																																			
2012	<p>2012 Keyword Distribution Data (Approximate):</p> <table border="1"> <tr><th>Keyword</th><th>Items</th></tr> <tr><td>FINANCIAL RISK</td><td>85</td></tr> <tr><td>DEBT CRISIS</td><td>80</td></tr> <tr><td>CURRENCY RISK</td><td>75</td></tr> <tr><td>FINANCIAL INSTRUMENT</td><td>70</td></tr> <tr><td>GLOBAL ECONOMY</td><td>65</td></tr> <tr><td>ECONOMIC GROWTH</td><td>60</td></tr> <tr><td>LIQUIDITY RISK</td><td>58</td></tr> <tr><td>LEGAL RISK</td><td>55</td></tr> <tr><td>GLOBAL ECONOMIC</td><td>52</td></tr> <tr><td>SOVEREIGN DEBT</td><td>50</td></tr> <tr><td>CAPITAL MARKET</td><td>48</td></tr> <tr><td>CREDIT RISK</td><td>45</td></tr> <tr><td>ECONOMIC DEVELOPMENT</td><td>42</td></tr> <tr><td>PERSONNEL RISK</td><td>40</td></tr> <tr><td>MARKET RISK</td><td>38</td></tr> </table>	Keyword	Items	FINANCIAL RISK	85	DEBT CRISIS	80	CURRENCY RISK	75	FINANCIAL INSTRUMENT	70	GLOBAL ECONOMY	65	ECONOMIC GROWTH	60	LIQUIDITY RISK	58	LEGAL RISK	55	GLOBAL ECONOMIC	52	SOVEREIGN DEBT	50	CAPITAL MARKET	48	CREDIT RISK	45	ECONOMIC DEVELOPMENT	42	PERSONNEL RISK	40	MARKET RISK	38	<p>2012 Proximity Plot Data (Approximate):</p> <table border="1"> <tr><th>Keyword</th><th>RISK</th></tr> <tr><td>FINANCIAL</td><td>0.90</td></tr> <tr><td>CONTROL</td><td>0.88</td></tr> <tr><td>MARKET</td><td>0.85</td></tr> <tr><td>DEVELOPMENT</td><td>0.82</td></tr> <tr><td>FINANCE</td><td>0.80</td></tr> <tr><td>OPERATION</td><td>0.78</td></tr> <tr><td>PRICE</td><td>0.75</td></tr> <tr><td>REGULARLY</td><td>0.72</td></tr> <tr><td>PERFORMANCE</td><td>0.70</td></tr> <tr><td>PRESENT</td><td>0.68</td></tr> <tr><td>COVER</td><td>0.65</td></tr> <tr><td>STRATEGY</td><td>0.62</td></tr> <tr><td>INVESTMENT</td><td>0.60</td></tr> <tr><td>SUBSIDIARY</td><td>0.58</td></tr> <tr><td>PLANNING</td><td>0.55</td></tr> </table>	Keyword	RISK	FINANCIAL	0.90	CONTROL	0.88	MARKET	0.85	DEVELOPMENT	0.82	FINANCE	0.80	OPERATION	0.78	PRICE	0.75	REGULARLY	0.72	PERFORMANCE	0.70	PRESENT	0.68	COVER	0.65	STRATEGY	0.62	INVESTMENT	0.60	SUBSIDIARY	0.58	PLANNING	0.55		
Keyword	Items																																																																			
FINANCIAL RISK	85																																																																			
DEBT CRISIS	80																																																																			
CURRENCY RISK	75																																																																			
FINANCIAL INSTRUMENT	70																																																																			
GLOBAL ECONOMY	65																																																																			
ECONOMIC GROWTH	60																																																																			
LIQUIDITY RISK	58																																																																			
LEGAL RISK	55																																																																			
GLOBAL ECONOMIC	52																																																																			
SOVEREIGN DEBT	50																																																																			
CAPITAL MARKET	48																																																																			
CREDIT RISK	45																																																																			
ECONOMIC DEVELOPMENT	42																																																																			
PERSONNEL RISK	40																																																																			
MARKET RISK	38																																																																			
Keyword	RISK																																																																			
FINANCIAL	0.90																																																																			
CONTROL	0.88																																																																			
MARKET	0.85																																																																			
DEVELOPMENT	0.82																																																																			
FINANCE	0.80																																																																			
OPERATION	0.78																																																																			
PRICE	0.75																																																																			
REGULARLY	0.72																																																																			
PERFORMANCE	0.70																																																																			
PRESENT	0.68																																																																			
COVER	0.65																																																																			
STRATEGY	0.62																																																																			
INVESTMENT	0.60																																																																			
SUBSIDIARY	0.58																																																																			
PLANNING	0.55																																																																			

Table 4. Keyword distribution and proximity plots per year

Firstly, it is necessary to address the statements of keyword distribution and proximity plots (see Table 4). A glance at the keyword distribution makes it clear that a comparable set of keywords takes precedence in all the observed years. Accordingly, for example, the keywords “financial risk”, “currency risks” or “financial instrument” were discussed regularly in a large portion of the risk reports (> 56% of cases). A closer inspection, however, makes it clear that there are also changes in the regularly discussed keywords from period to period. In 2006, for example, “economic growth” (42%) was regularly discussed, while in 2009 terms such as “economic crisis” (68%), “financial crisis” (52%), “economic downturn” (44%), or “credit risk” (52%) are at the center of the general discussion. Therefore, the financial crisis and its consequences appear to have made their way into the reporting. In 2012, however, these keywords disappeared from the discussion, and the regular appearance of “economic growth” (62%) returned again.

The influence of the financial crisis can also be seen in the results of the proximity plots. In particular, the aspects of market prices in connection with the term “risk” were frequently discussed in 2006 (i.a., “interest”, “hedge”, “currency”, “exchange”). By contrast, in 2009, credit risks were increasingly in focus (i.a., “investment”, “capital”, “liability”). In 2012, strategic (“strategy”, “planning”, “subsidiary”) and operative aspects (“operation”, “present”, “performance”) were mentioned in connection with the term “risk”. In summary, these exploratory findings clearly indicate that some significant changes in the risk discussion took place, thus representing a good starting point for further explorations such as the following topic extraction or link analysis.

The topical concepts uncover the hidden thematic structure in the risk discussions (see Table 5). All in all, five different topics were extracted per year. If one compares the topics in 2006, 2009, and 2012, multiple similarities in the risk discussion become clear. Accordingly, for example, aspects of hedging and the risk reduction of price changes (e.g., exchange rates) are regularly discussed (topic no. 5 in 2006; no. 4 in 2009; no. 3 in 2012). Likewise, market price-oriented risks (combined discussion of “price”, “market” and, e.g., “material”, “energy”, “services”) were a regular topic (topic no. 1 in 2006; no. 2 in 2009; no. 5 in 2012).

Year	Topics						
	NO	TOPIC	KEYWORDS	% VAR	FREQ	CASES	% CASES
2006	1	PRICE; MATERIAL	MATERIAL; PRICE; COST	6.56	547	38	100.00%
	2	CUSTOMER; SALES AND SERVICE	CUSTOMER; SERVICE; QUALITY; COST; SALE	5.60	802	38	100.00%
	3	GROWTH MARKET	GROWTH; ECONOMIC; GLOBAL; MARKET	4.80	1246	38	100.00%
	4	FINANCIAL RISK	RISK; CONTROL; LIMIT; FINANCIAL; CREDIT	4.64	1483	38	100.00%
	5	HEDGE; CURRENCY EXCHANGE	HEDGE; CURRENCY; TRANSACTION; EXCHANGE; INTEREST	4.58	678	37	97.37%
2009	1	ECONOMIC CRISIS; GLOBAL ECONOMY	ECONOMIC; GLOBAL; ECONOMY; CRISIS; RECOVERY; MARKET; GROWTH	2.06	2469	38	100.00%
	2	MARKET; PRICE	PRICE; COST; CUSTOMER; MATERIAL; MARKET; SERVICE; ENERGY	1.85	1521	38	100.00%
	3	CREDIT; LIQUIDITY	CREDIT; LIQUIDITY; CASH; BANK	1.96	589	38	100.00%
	4	HEDGE; CURRENCY	HEDGE; CURRENCY; TRANSACTION; INTEREST; EXCHANGE	1.96	853	38	100.00%
	5	COMPLIANCE;LAW; REGULATION	LAW; COMPLIANCE; REGULATION; ENVIRONMENTAL; PRINCIPLE; GUIDELINE; RULE	1.76	846	37	97.37%

	NO	TOPIC	KEYWORDS	% VAR	FREQ	CASES	% CASES
2012	1	COMPLIANCE; REGULATION	COMPLIANCE; REGULATION; GUIDELINE; POLICY LAW; RULE;	1.99	1111	38	100.00%
	2	CONTROL AND RISK	RISK; PLANNING CONTROL;	1.81	2380	38	100.00%
	3	HEDGE; FINANCIAL	HEDGE; INSTRUMENT; TRANSACTION; EXCHANGE; FINANCIAL CURRENCY; INTEREST;	1.84	2117	38	100.00%
	4	GROWTH; MARKET	ECONOMY; ECONOMIC; EMERGE; MARKET GLOBAL; GROWTH;	1.83	3055	38	100.00%
	5	MARKET; PRICE	PRICE; PRODUCTION; CUSTOMER; MATERIAL MARKET; COST;	1.74	1518	38	100.00%

Table 5. Topics per year

When considered in more detail, it also becomes clear, however, that there are differences in the risk discussion from period to period. In 2009, for example, the influence of the crisis on global economic growth is a subject for the first time (topic no. 1). A discussion around the terms “global”, “economy”, “crisis”, “recovery”, “market”, and “growth” form a central topic in the risk reporting. This again shows that the economic impact of the global financial crisis made its way into the risk discussion of the German industry. In 2012, however, the crisis seems to have disappeared from the focus of the discussion. In 2012, regulatory aspects form a central topic in the risk reporting (topic no.1 contains a discussion of aspects such as “compliance”, “regulation”, and “law”). This might possibly indicate a changed legal situation in the German industry. Attention should be paid not only to what is disclosed and its volume but also to how it is disclosed, which means in what context it is discussed. For this purpose, the link analysis complements the previous findings by providing visuals for the relationship patterns of the connected topics. As noted above, the economic crisis appeared in the risk discussion for the first time in 2009. A link analysis is applied in order to discover and visualize more detailed content relationships with respect exclusively to the term “crisis”. Figure 3 displays the three associated concepts with this term.

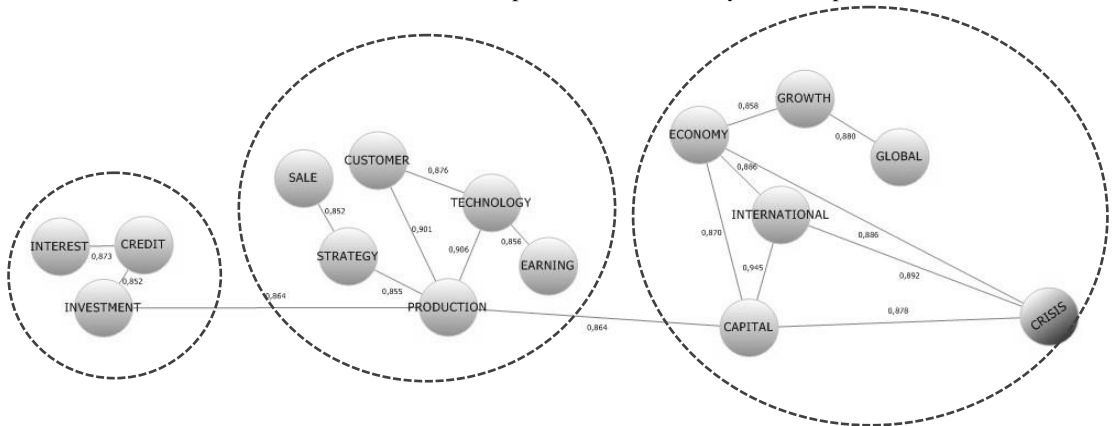


Figure 3. Link analysis.

First, a discussion of economic issues becomes apparent (a combined discussion of “economy”, “international”, “growth”, global” and “capital”). Second, a related discussion of the production can be identified (i.e., “capital” in connection with “production”, “customer”, “sale”, “strategy”, “sale”), which indicates connected problems with regard to declines in sales. Third, an associated discussion related to investments is identifiable (“investment”, credit”, interest”), which indicates involved investment difficulties. Altogether, these exploratory findings very aptly illustrate the effects of the economic crisis on the industrial sector.

5. DISCUSSION OF APPLICATION EXPERIENCES

The experiences gained from the application of the two text analysis approaches shall now be summarized and discussed. In general, the following advantages of computer-aided processes could be identified:

- The automated collection of information in many cases, e.g., via web crawler functions (automated collection of content published online), can make it substantially easier to create textual databases.
- The preparation of the 114 annual reports in a structured database (TDM) took place mostly automatically. Nevertheless, it should be noted that the evaluation and adjustment of this database (e.g., elimination of irrelevant terms and correction of misspellings) took some considerable time.
- The use of computer-assisted processes proved to be an efficient solution both of automating the coding process to the greatest extent possible and

of processing large data sources as well as a large number of keywords. In this way, far more documents can be processed than would be possible in manual analyses.

- The identification and counting of the predefined keywords is repeatable and reliable since it takes place independently of human experience, subjective interpretation or attention.
- The results of the word counts and the statistical analyses and tests are available to analysts automatically. The repeatability of analyses and their results has proven advantageous since the corrections for subsequent ideas or changes (e.g., replacement or correction of keywords) could be adjusted and repeated at any time.
- The computer-assisted processes facilitate exploratory analyses (e.g., identification of content relationships and patterns), which cannot be conducted in the framework of manual analyses or can only be done with a great effort.

In general, the two observed analytical approaches have proven to be compatible and mutually complementary. On the one hand, the quantitative results of the dictionary approach are particularly well-suited to examine the volume, dissemination and trend of corporate disclosures; on the other hand, the statistical approach could reveal new insight into the discussions and their content relationships. Nonetheless, methodological risks and limitations in the respective approaches also became clear. These shall be discussed in more detail in the following.

5.1 Dictionary Approach

The dictionary approach has proven to be an effective solution for producing precise quantitative results about the volume, dissemination and development of reported concepts. After a dictionary with thematic categories and keywords has been modelled, it is possible to use computer-assisted means to code large texts very quickly and to produce quantitative, statistically verifiable statements on the thematic categories. The advantages of computer-assisted processes, however, also come with limitations. Firstly, the dictionary approach is highly dependent on the validity, objectivity, and completeness of the underlying thematic categories

and keywords (Krippendorff, 2013). On the one hand, the selection of the keywords themselves has a substantial influence on the results of this analysis. Other keywords or another combination can certainly reveal different results. On the other hand, the keywords identified in texts may appear in a completely different context than the one envisaged in the dictionary. For the securing of meaningful and unambiguous dictionaries, it is therefore recommended that “sources of certainty” be relied on, such as expert knowledge, existing theories, or practically tested concepts (Krippendorff, 2013). Furthermore, checking the plausibility of the keywords (by a keyword-in-context check or verification by third parties) can ensure the meaningfulness and absence of overlapping with other words. Secondly, the dictionary approach is strongly based on the assumption that the frequency of the identified keywords is a direct indicator for their relevance (Weber, 1990). Qualitative evaluations, particularly in connection with the intensity or the meaningfulness of the described content, are not taken into account in such automated processes. Accordingly, the exact context of the words is largely left out in this (e.g., are credit risks gaining or losing importance, and if they are, why?). Thirdly, the discovery of new phenomena is not possible through this predefined and limited focus. The dictionary approach limits the analysis by a predefined framework and does not provide any possibility to extract previously unknown concepts from the material (Indulska *et al.*, 2012).

5.2 Statistical Approach

The findings exemplify how statistical analyses can uncover previously unknown content and relationships in the risk discussion, and therefore should have better chances of producing early warnings of beginning crises. Such statistical approaches explore the texts without limitations and, therefore, make it possible to overcome the fixed framework of the dictionary approach. For this purpose, diverse statistical techniques and visualizations are at the analyst's disposal. Nonetheless, the results of such exploratory processes can often only be seen as a starting point for further and more detailed analyses. The insight derived from link analyses, for example, hold great potential for uncovering new relationships but, however, they can hardly be seen as a certain and unambiguous proof. Further qualitative analyses should verify the indicated content relationships and extent the knowledge about them. Furthermore, because of their exploratory nature, such

uncontrolled analyses make the securing of valid results, meaning their interpretation and evaluation, much more difficult. Krippendorff (2013, p. 35) described it as follows: “sophisticated statistical techniques will always yield something, but this does not guarantee that the results will refer to anything.” Although statistical text analyses have proven to be efficient and promising solutions for the extraction of new knowledge, analysts should not trust this process blindly. Statistical text analyses require a manual effort that cannot be underestimated, particularly with regard to the interpretation and plausibility check of the generated statements. This means, on the one hand, that the clear connection to the initially defined questions should always remain and, on the other, that the actual core statements of the texts may not be distorted. Krippendorff (2013, p. 212) stressed this demand as follows: “To justify the use of [text analysis] software, content analysts must assure themselves, as well as the community of peers, that the way a software package processes the data is compatible with what is known about the context of the texts, how texts are read, what they mean, and what role they play.”

6. SUMMARY AND CONCLUSIONS

This paper pursued the goal of demonstrating the use of computer-aided text analysis approaches (dictionary and statistical approach) and supplying the accounting discipline with recommendations for corresponding analyses of textual corporate disclosures. On the one hand, computer-assisted processes have proven to be a very efficient solution for coping with large text collections, allowing the exploration of large disclosure volumes and thus helping to overcome the problem of information overload. On the other hand, the combined use of both text analysis approaches has proven advantageous since they complement each other and compensate for each other's weaknesses. The combination of quantitative results related to thematic categories (dictionary approach) as well as the exploration of unknown content and relationships (statistical approach) created a more comprehensive picture with regard to the presentation of corporate disclosure. By this means, it can be explained not only what is disclosed and how much but also in what way.

In the light of the practical experiences collected during this case study, several recommendations for implementing comparable computerized text analyses can

be made:

- A computer-aided text analysis requires consistent planning and a detailed documentation of all decisions and technical settings (especially in the context of data preparation and statistical processes). For this purpose, the consideration of the introduced text analysis process can be recommended.
- The processes of data selection and preparation should include only relevant textual content. In this way, the risk of statistical distortions resulting from irrelevant data can be avoided. Furthermore, the effort for manual adjustments and corrections of the resulting database should not be underestimated.
- The common software tools offer several functions for modelling precise and machine processable dictionaries. Nevertheless, in order to safeguard the objectivity as well as validity of thematic dictionaries, the consideration of profound “sources of certainty” such as established theories, expert knowledge and experiences, or embodied practices is recommended. Furthermore, it is recommended to check the meaning and suitability and of keywords by checking random samples in their respective context (keyword-in-context) or to verify them by a third party.
- Exploratory text analysis approaches require a careful interpretation and plausibility check in order to assure valid insight resulting from automated processes. This effort cannot be underestimated. Furthermore, they can rarely be seen as certain proof and further qualitative analyses should verify the indicated content relationships and extent the knowledge about them.

There are two avenues for further research. First, other potential areas of application could be evaluated. Future studies could use computer-aided approaches in order to investigate other issues on the basis of corporate disclosures, such as, for example, the strategies and actions of competitors (in the sense of a competitor analysis). Another interesting topic is the automated prediction of financial performances on the basis of corporate disclosures (Kloptchenko *et al.*, 2004). Second, this demonstration of computer-aided text analyses could not cover the full technical potential of these software tools.

Common text analysis programs offer a variety of analytic techniques and visualizations (Feldman & Sanger, 2007; Miner *et al.*, 2012; Weiss *et al.*, 2010). Therefore, future studies could extend the methodological knowledge by demonstrating and discussing further analytic applications (e.g., document classification or clustering techniques).

This study is not without limitations. First, the exploratory case study has been conducted by the researchers themselves. This may have introduced a certain subjective interpretation into the discussion. It would therefore be helpful to verify the recommendations in practice. Second, the main focus of this study lied on the demonstration of computerized analyses and rather less on the empirical study itself. That is why the results have been discussed only briefly. Nonetheless, corresponding explorations of the risk reporting seem promising. Further contributions could continue this analysis and extend the findings obtained from this study.

7. REFERENCES

- ALWERT, K.; BORNEMANN, M.; WILL, M. (2009): "Does intellectual capital reporting matter to financial analysts?", *Journal of Intellectual Capital*, vol. 10, n. 3: 354-368. <http://dx.doi.org/10.1108/14691930910977789>
- APCA (Association of Public Certified Accountants) (2010). Hitting the notes, but what's the tune? An international survey of CFO's views on narrative reporting. Retrieved October 18, 2013, from: <http://goo.gl/TtpgSw>
- AICPA (1994): Improving Business Reporting - A Customer Focus: Meeting the Information Needs of Investors and Creditors. *Report of the Special Committee on Financial Reporting*. American Institute of Certified Public Accountants, New York.
- BEATTIE, V.; THOMSON, S. J. (2007): "Lifting the lid on the use of content analysis to investigate intellectual capital disclosures", *Accounting Forum*, vol. 31, n. 2: 129-163. <http://dx.doi.org/10.1016/j.accfor.2007.02.001>
- BENBASAT, I.; GOLDSTEIN, D. K.; MEAD, M. (1987): "The case research strategy in studies of information systems", *MIS quarterly*, vol. 11, n. 3: 369-386.

<http://dx.doi.org/10.2307/248684>

BEYNON, M. J.; CLATWORTHY, M. A.; JONES, M. J. (2004): "The prediction of profitability using accounting narratives: a variable-precision rough set approach", *Intelligent Systems in Accounting, Finance and Management*, vol. 12, n. 4: 227-242. <http://dx.doi.org/10.1002/isaf.256>

BOESSO, G.; KUMAR, K. (2007): "Drivers of corporate voluntary disclosure: A framework and empirical evidence from Italy and the United States", *Accounting, Auditing & Accountability Journal*, vol. 20, n.2: 269-296.

BOTOSAN, C. A. (1997): "Disclosure level and the cost of equity capital", *The Accounting Review*, vol. 72, n. 3: 323-349. <http://dx.doi.org/10.2139/ssrn.224385>

BOWMAN, E. H. (1984): "Content analysis of annual reports for corporate strategy and risk", *Interfaces*, vol. 14, n. 1: 61-71.

CAS (Casualty Actuarial Society) (2003): Overview of Enterprise Risk Management. Enterprise Risk Management Committee. Retrieved October 18, 2013, from: <http://www.casact.org/area/erm/overview.pdf>

CAMPBELL, D.; SHRIVES, P.; BOHMBACH-SAAGER, H. (2001): "Voluntary disclosure of mission statements in corporate annual reports: signaling what and to whom?", *Business and society review*, vol. 106, n.1: 65-87. <http://dx.doi.org/10.1111/0045-3609.00102>

DIRSMITH, W.; COVALESKI, A. (1983): "Strategy, external communication, and environmental context", *Strategic Management Journal*, vol. 4, n. 2: 137-151.

DURST, S. (2013): "An exploratory study of intangibles risk disclosure in annual reports of banking companies from the UK, US, Germany and Italy - Some descriptive insights", *Financial Reporting*, vol. 1, n. 1: 81-120. <http://dx.doi.org/10.3280/FR2013-001005>

FEI (Financial Executives International) (2001): Strengthening Financial Markets: Do Investors have the Information they need?, Retrieved May 22, 2013, from: www.fei.org/finrep/files/SEC-Taskforce-Final-6-6-2k1.pdf

FELDMAN, R.; SANGER, J. (2007): "*The text mining handbook: advanced approaches in analyzing unstructured data*". Cambridge, MA: University Press.

FASB (2001): *Improving Business Reporting: Insights into Enhancing Voluntary Disclosures*. New York, NY: Business Reporting Research Project. Steering Committee.

FRANCIS, J.; NANDA, D.; OLSSON, P. (2008): "Voluntary disclosure, earnings quality, and cost of capital", *Journal of Accounting Research*, vol. 46, n. 1: 53-99. <http://dx.doi.org/10.1111/j.1475-679X.2008.00267.x>

FTSE (2013): *Industry Classification Benchmark (ICB)*. Retrieved October 14, 2013, from: <http://www.icbenchmark.com/http://www.icbenchmark.com/>

GAMERSCHLAG, R.; MÖLLER, K.; VERBEETEN, F. (2011): "Determinants of voluntary CSR disclosure: empirical evidence from Germany", *Review of Managerial Science*, vol. 5, n. 2-3: 233-262. <http://dx.doi.org/10.1007/s11846-010-0052-3>

GAO, L.; CHANG, E.; HAN, S. (2005): "Powerful tool to expand business intelligence: Text mining", *Proceedings of World Academy of Science, Engineering and Technology*, vol. 8: 110-115.

GASB (German Accounting Standards Board) (2012): *German Accounting Standards (GAS)*. Berlin.

GOEL, S.; GANGOLLY, J. (2012): "Beyond the numbers: Mining the annual reports for hidden cues indicative of financial statement fraud", *Intelligent Systems in Accounting, Finance and Management*, vol. 19, n. 2: 75-89. <http://dx.doi.org/10.1002/isaf.1326>

GUTHRIE, J.; PETTY, R.; YONGYANICH, K.; RICCERI, F. (2004): "Using content analysis as a research method to inquire into intellectual capital reporting", *Journal of intellectual capital*, vol. 5, n. 2: 282-293. <http://dx.doi.org/10.1108/14691930410533704>

HEALY, P. M.; PALEPU, K. G. (2001): "Information asymmetry, corporate disclosure, and the capital markets: A review of the empirical disclosure literature", *Journal of accounting and economics*, vol. 31, n. 1: 405-440. <http://dx.doi.org/10.2139/ssrn.258514>

HOOKE, J.; COY, D.M.; DAVEY, H. (2002): "The information gap in annual reports", *Accounting, Auditing & Accountability Journal*, vol. 15, n. 4: 501-522.

<http://dx.doi.org/10.1108/09513570210440577>

INDULSKA, M.; HOVORKA, D. S.; RECKER, J. (2012): "Quantitative approaches to content analysis: identifying conceptual drift across publication outlets", *European Journal of Information Systems*, vol. 21, n. 1: 49-69.

IASB (International Accounting Standards Board) (2010): Management Commentary - A framework for presentation. IFRS Practice Statement. Retrieved October 18, 2013, from: <http://www.ifrs.org/Current-Projects/IASB-Projects/Management-Commentary/IFRS-Practice-Statement/Pages/IFRS-Practice-Statement.aspx>

KLOPTCHENKO, A.; MAGNUSSON, C.; BACK, B.; VISA, A.; VAHARANTA, H. (2004): "Mining textual contents of financial report", *The International Journal of Digital Accounting Research*, vol. 4, n. 7: 1-29.

KRIPPENDORFF, K. (2013): "*Content Analysis. An Introduction to its Methodology*". 3rd Edition. Thousand Oaks, CA: Sage Publications.

LI, F. (2010): "Textual analysis of corporate disclosures: A survey of the literature", *Journal of Accounting Literature*, vol. 29: 143-165.

MANNING, C. D.; SCHÜTZE, H. (1999): "*Foundations of statistical natural language processing*". Cambridge, MA: MIT Press.

MCNIFF, J.; WHITEHEAD, J. (2000); "*Action Research in Organisations*". London: Routledge.

MINER, G.; ELDER, J.; HILL, T.; NISBET, R.; DELEN, D.; FAST, A. (2012): "*Practical Text Mining and Statistical Analysis for Non-Structured Text Data Applications*". 1st Edition. Cambridge, MA: Academic Press.

MORRIS, R. (1994): "Computerized content analysis in management research: A demonstration of advantages & limitations", *Journal of Management*, vol. 20, n. 4: 903-931. <http://dx.doi.org/10.1177/014920639402000410>

OECD (Organisation for Economic Co-operation and Development) (2012): Corporate Reporting of Intangible Assets: A Progress Report. Retrieved October 14, 2013, from: <http://www.oecd.org/daf/ca/corporatereportingofintangibleassetsaprogressreport.htm>

- PROBOHUDONO, A. N.; TOWER, G.; RUSMIN, R. (2013): "Risk disclosure during the global financial crisis", *Social Responsibility Journal*, vol. 9, n. 1: 124-137. <http://dx.doi.org/10.1108/17471111311307859>
- SARLIN, P.; MARGHESCU, D. (2011): "Visual predictions of currency crises using self-organizing maps", *Intelligent Systems in Accounting, Finance and Management*, vol. 18, n. 1: 15-38. <http://dx.doi.org/10.1002/isaf.321>
- SCHLÜTER, T.; HARTMANN-WENDELS, T.; WEBER, T.; ZANDER, M. (2014): "Risk Disclosure of German Banks: Establishing the Industry Standard", (August/September): 386-427.
- SHIRATA, C. Y.; TAKEUCHI, H.; OGINO, S.; WATANABE, H. (2011): "Extracting key phrases as predictors of corporate bankruptcy: Empirical analysis of annual reports by text mining", *Journal of Emerging Technologies in Accounting*, vol. 8, n. 1: 31-44. doi: <http://dx.doi.org/10.2308/jeta-10182>
- SUBRAMANIAN, R.; ISHAK, S. T. (1998): "Competitor analysis practices of US companies: An empirical investigation", *MIR: Management International Review*, vol. 38, n. 1: 7-23.
- WEBER, R.P. (1990): "*Basic Content Analysis*". 2nd Edition. Newbury Park, CA: Sage Publications.
- WEISS, S. M.; INDURKHYA, N.; ZHANG, T.; DAMERAU, F. (2010): "*Text mining: predictive methods for analyzing unstructured information*". New York, NY: Springer Science & Business Media.
- WGARIA (working group "Accounting and Reporting of Intangible Assets" of the Schmalenbach Society) (2005): "Corporate Reporting on Intangibles – A Proposal from a German Background". *Schmalenbach Business Review*, Special Issue 2/05: 65-100.