RCC Revista Contemporânea de Contabilidade

Do explanatory notes explain? Risk management disclosure analysis using text mining

Notas explicativas explicam? Análise da comunicação do gerenciamento de risco a partir de técnicas de text mining

¿Notas explicativas explican? Análisis de la comunicación de la gestión de riesgos a partir de técnicas de text mining

Carla Bonato Marcolin*
Doutora em Administração (UFRGS) e Pós-Doutoranda na Escola de Administração de Empresas de São Paulo (FGV)
Professora da Faculdade de Gestão e Negócios (UFU), Uberlândia/MG, Brasil
cbmarcolin@gmail.com
https://orcid.org/0000-0003-0260-5073

Enrico Dalla Riva
Mestre em Controladoria e Contabilidade (USP) e Doutorando na Escola de Administração de Empresas de São Paulo (FGV), São Paulo/SP, Brasil
enrico.riva@bcb.gov.br
https://orcid.org/0000-0001-7032-6039

Fernanda da Silva Momo
Doutora em Administração (UFRGS)
Professora da Faculdade de Ciências Econômicas (UFRGS), Porto Alegre/RS, Brasil
fernanda.smomo@gmail.com
http://orcid.org/0000-0002-6512-5280

João Luiz Becker
Doutor em Management Science na University of California at Los Angeles (UCLA)
Professor Titular de Tecnologia e Ciência de Dados na EAESP (FGV), São Paulo/SP, Brasil
beckerjoaoluiz@gmail.com
https://orcid.org/0000-0003-4176-7374

Primary contact address for correspondence*
Av. João Naves de Ávila, 2121, Bairro Santa Mônica, CEP: 38408-144 – Uberlândia/MG, Brasil

Abstract
The study aimed to analyze the effectiveness of risk management communication in the explanatory notes considering text mining techniques. Based on the analysis of 241 text excerpts from 32 financial institutions, we used the cosine similarity as a distance measurement to assess the relationship between the variability of the texts and the Basel and Inad 90 risk indices. The results indicate that the analyzed notes are not in compliance with OCPC 07, as their content does not reflect the fluctuation of risk indices. The benefit of the notes for the user in decision-making processes may be impaired, given the evidence of disconnection between the texts and the risk policy. Cases of complete repetition of the text excerpts are presented. As a contribution, this article advances in the textual studies of financial statements, going beyond lexical and volumetric analyses, bringing the semantic content in volume and its relationship with finance indicators.
Keywords: Explanatory notes; Text mining; Risk management

Resumo
O estudo objetivou analisar a efetividade da comunicação do gerenciamento de risco nas notas explicativas considerando técnicas de *text mining*. A partir da análise de 241 trechos de texto de 32 instituições financeiras, utilizou-se a similaridade de cosseno como medida de distância para avaliar a relação da variabilidade dos textos com os índices de risco Basileia e Inad90. Os resultados indicam que as notas analisadas não estão em conformidade com a OCPC 07, pois seu conteúdo não reflete a flutuação dos índices de risco. A utilidade para o usuário em processos de tomada de decisões pode ser prejudicada, dado o indício de desconexão entre os textos e a política de riscos. Apresentam-se casos de repetição integral dos trechos de texto. Como contribuição, esse artigo avança nos estudos textuais das demonstrações financeiras, indo além de análises léxicas e volumétricas, trazendo o conteúdo semântico em volume e sua relação com indicadores financeiros.
Palavras-Chave: Notas explicativas; *Text mining*; Gerenciamento de risco

Resumen
El estudio tuvo como objetivo analizar la efectividad de la comunicación de la gestión de riesgos en las notas explicativas considerando técnicas de minería de textos. A partir del análisis de 241 extractos de texto

de 32 instituciones financieras, se utilizó la similitud del coseno como medida de distancia para evaluar la relación entre la variabilidad de los textos y los índices de riesgo de Basilea e Inad90. Los resultados indican que las notas analizadas no están en conformidad con OCPC 07, ya que su contenido no refleja la fluctuación de los índices de riesgo. La utilidad para el usuario en los procesos de toma de decisiones puede verse afectada, dada la evidencia de desconexión entre los textos y la política de riesgos. Se presentan casos de repetición completa de extractos del texto. Como contribución, este artículo avanza en los estudios textuales de los estados financieros, yendo más allá de los análisis léxicos y volumétricos, trayendo el contenido semántico en volumen y su relación con los indicadores financieros.
Palabras-Clave: Notas explicativas; *Text mining*; Gestión de riesgos

1 Introduction

Computational advances allow artificial intelligence tools to be more accessible in organizational environments. Increasing processing and memory capacities expand the possibilities of analysis, bringing other data formats previously unavailable. At the same time, the Big Data scenario brings in one of its dimensions: V, for Variety, comprising unstructured data (McAfee et al., 2012), which is the object of analysis in this article. Unstructured data is any data whose format does not allow traditional organization in tables or relational databases. Examples are social media data such as videos and photos, audio data, and text data. This last format, textual, has been the focus of several pieces of research in the accounting-financial area due to its flexibility and information value, which can generate valuable insights for decision-making processes of different natures (Fisher et al., 2016; Sawanobori, 2013; Yang et al., 2018).

Text mining techniques permit extracting, summarizing, and classifying texts, allowing their manipulation through quantitative procedures. In this article, the objective is to analyze the effectiveness of the risk management disclosure of financial institutions from the content of the explanatory notes, considering text mining techniques. Thus, we intend to explore the possibility of showing the effectiveness of risk management disclosure in financial institutions based on an analysis whose constitutive source is the textual content of the explanatory notes. Explanatory notes are the space for detailing the financial statements presenting, among others: the basis for the preparation of the reports and the specific policies used; complementary information required by regulations; additional information relevant to the understanding of the organization's financial information (CPC, 2011; IFRS, 2020).

Given the relevance of the content presented in the explanatory notes and the number of textual fragments existing in them, there is a growing search for the application of data analytics techniques that allow better use of this information for decision making, improvement of audits, and how to present the information to its users. Several studies use data analytics techniques to improve estimates and writing, analyze managers' speeches, predict fraudulent actions, among others (Ding et al., 2019; Loughren & McDonald, 2016). This study examines the communication on the risk positioning present in the explanatory notes. It aims to identify whether stakeholders can utilize this source of text as a support for decision making concerning financial institutions, or if the explanatory notes do not yet represent faithfully the positioning of institutions about their risk management practices, which imply that the explanatory notes would not be explaining what they are supposed to explain.

There is a discussion about using the explanatory notes to promote accountability, faithfully representing business practices regarding risk management, or whether firms use them merely to comply with regulations, without concern to explain and communicate to the stakeholders the organization's positions on the topic (Hendriksen & Van Breda, 2010; Nakagawa et al., 2008). Therefore, using explanatory notes for risk analysis can provide a new view of the risk area for control institutions and investors.

After this introduction, the following section presents a theoretical framework on accounting disclosure, risk, and explanatory notes, displaying the research gap. Section three addresses some aspects of methodological choices, followed by the section where the results are analyzed. Finally, section five presents the conclusions, limitations, and next steps of this study.

2 Theoretical Framework

Below are three sections covering the accounting disclosure process (2.1), the information disclosed on risk management in financial institutions (2.2), and the explanatory notes (2.3).

2.1 Accounting Disclosure

The accounting information disclosure process finds theoretical support in the motivations for the adoption of transparency standards in organizations. Verrecchia (2001) suggests three categories of studies of incentive factors for accounting disclosure: (a) based on association, supported by studies on the effects of information on investors decision making; (b) based on judgment, due to the managers' power to disclose or omit internal information; and (c) based on efficiency, which indicates that any information is better than

the absence of it. Verrecchia (2001) suggests the development of a theory that involves all these factors, based on the process of reducing informational asymmetries.

An example of the first approach is the relationship between the disclosure of accounting information and the change in the value of the companies' shares. The second approach seeks to investigate the incentives of managers or companies to promote disclosure, such as a possible decrease in the company's value when the company does not disclose relevant information. Finally, in the efficiency-based approach, firms seek to reduce informational asymmetry to reduce this factor in their cost of capital (Yamamoto & Salotti, 2006).

Dantas et al. (2005) investigated the duality between the benefits of disclosure and the reluctance of organizations to disclose internal information to the public. While disclosing information improves the value of entities and the reliability of creditors and investors, there is an aversion to increasing the degree of disclosure with the argument of protecting strategic information. Concerning financial institutions, the authors consider that the demand for appropriate disclosure is even more necessary to inform users of the risks to which the institution is exposed (Dantas et al., 2005).

The phenomenon of accounting disclosure to the market may have voluntary or mandatory characteristics. Proponents of compulsory disclosure consider that entities do not increase the level of revelation without legal imposition (Yamamoto & Salotti, 2006). The present study considers aspects of Verrecchia's second proposition (Verrecchia, 2001), based on the managers' discretion regarding disclosing accounting information in the face of an obligation to provide information defined by the regulator.

2.2 Information on Risk Management

Financial institutions, in particular, have at the core of their operations the presence of risk factors and uncertainties inherent to banking businesses. Therefore, the risk management process is of critical importance, as a clear understanding of the risks incurred in the operational routines is necessary to assure proper confront, control, and management of risks (Raghavan, 2003). According to Crockford (1982), the effectiveness of risk management relates directly to successful loss control.

Some international initiatives have emerged to promote good corporate practices in internal procedures, thus becoming a reference. The COSO (Committee of Sponsoring Organizations of the Treadway Commission), which advises the Security Exchange Commission (SEC) in the United States, was one of those organizations that produced material relevant to the topic (Benjamin Jr., 2005). Enterprise risk management is a process, applied in strategy setting and across the enterprise, designed to identify potential events that may affect the entity and manage risk to be within its risk appetite, to provide reasonable assurance regarding the achievement of entity objectives (COSO, 2004).

The international benchmark on financial systems is the Basel Committee on Banking Supervision, a part of the Bank for International Settlements (BIS), which recommends standards for capital and risk management approaches aimed at banking regulators worldwide. In its Core Principle 7, the entity suggests that supervisors of banking systems must ensure that supervised institutions have a comprehensive risk management process that includes identification, assessment, monitoring, control and mitigation, and capital adequacy to their risk profile (BCBS, 1997).

In Brazil, the Central Bank of Brazil (Banco Central do Brasil - BCB), which follows the guidelines issued by the Basel Committee, regulates Brazilian financial institutions. Currently, its Resolution 4,557/17 presents the requirements in the regulation on risk management, which demands financial institutions to have an adequate integrated risk management structure compatible with their activities. The resolution article 56 requires the risk and capital management structure disclosure in a publicly accessible report and a summary of the description in the explanatory notes accompanying the half-yearly financial statements (BCB, 2017).

Torres and Galdi (2013) have already explored the level of adherence to the disclosure of information produced by financial institutions on risk management to the standards issued by the regulator in Brazil and noticed significant possibilities for improvement in the levels of reports. Ryan (2012) identified that the risk information released by the American financial institutions was increasingly extensive and with opaque integration with the financial statements, making it difficult for the users to identify their relevance. His work suggests that regulators should improve the quality of information requirements by standardizing the disclosure of risk factors that allow comparability, conducting stress tests with evidence of results, and closing ties with accounting information.

The literature shows evidence that the disclosure of risk management information of financial institutions has a positive relationship with the valuation of bank shares (Jizi & Dixon, 2017), financial results (Oino, 2019), stability (Del Gaudio et al., 2020), and governance (Nahar et al., 2020). In addition, the level of risk disclosure has predictive power over the risk rating of financial institutions (Elamer et al., 2021), allowing to nullify uncertainties (Kim et al., 2019). Even the risk management process disclosure is an essential part of the implementing risk culture and governance process in banking institutions (Schmitt, 2017).

The deepening of the analysis of financial reports using text mining techniques becomes relevant due to the volume of information generated in these reports, making it difficult to identify patterns, evidence

of fraudulent actions, or the quality of the information when the data is analyzed separately and manually. Text mining expands research examining the narratives of financial reports (Brown et al., 2020; Dyer et al., 2017; Huang et al., 2018).

2.3 Explanatory Notes

The explanatory notes, according to Technical Pronouncement 26 - Presentation of Financial Statements (CPC 26 (R1)), "offer narrative descriptions or breakdown (detail) of items presented in statements and information about items that do not qualify to be recognized in the financial statements" (CPC, 2011), which highlights the complementary but distinctive character of the explanatory notes. Likewise, the International Financial Reporting Standards (IFRS) highlight the explanatory notes as relevant elements for the existence of a complete financial report since they present the adopted policies, in addition to other information relevant to the public (IFRS, 2020). Highlighting the importance of the information included in the explanatory notes as a source of transparency and accountability.

After identifying explanatory notes presenting irrelevant information while lacking relevant information, the Brazilian Committee of Accounting Statements (CPC) issued technical guidance OCPC 07. It concluded that guidelines for disclosure of explanatory notes already existed, both in CPC 26 and in the Basic Conceptual Pronouncement - Conceptual Framework for Elaboration and Disclosure of Accounting and Financial Report (CPC, 2014), normative received by the Central Bank of Brazil by Resolution 4,144/12 (BCB, 2012), as the regulator of the financial system. The OCPC 07 highlights the importance of said accounting piece in providing material and relevant information that is useful to users of information when making decisions about entities, emphasizing the need to avoid the excess of irrelevant information that may deviate from the objective of informing, thus confusing the user (CPC, 2014). In this regard, a process of institutionalizing specific explanatory notes through regulation has already been identified in Brazil to increase the capacity to control and inspect risk markets (Santos et al., 2010).

Unlike quantitative indices, text data has a greater subjectivity and variability, not only in format but also in content, which means that they have information with a different substance. This specificity ends up creating an analysis barrier in an area dominated by financial quantifications, which often makes this type of data invisible (Fisher et al., 2010; Porter, 2005).

The advancement of text mining techniques has helped to change this scenario. Fisher et al. (2010) identified that the accounting area has an interface with the Natural Language Processing area (NLP) in two directions. The first relates with Information Retrieval (IR), trying to understand the lexicon of accounting and forming dictionaries (Nielsen & Fuertes-Oliveira, 2013; Swandi et al., 2017).

Although there are several studies in this vein, there are criticisms that point to the lack of quantitative semantic analysis and its volume, reducing the ability to integrate this data with other accounting information (Gomes et al., 2018). This criticism leads to the second direction, which is the focus of this study, connecting the data in accounting text to the text mining area. Both currents contribute to greater accessibility and understanding of these texts. This latter direction seeks to use text to carry out current analyzes and inferences, such as share prices and performance and fraud predictions (Loughran & McDonald, 2016; Rahrovi et al., 2019). In this way, more than creating a specific lexicon or linguistic set of accounting, text mining techniques can help to make this data more present in accounting assessments for managers and stakeholders.

However, even though there is a mature area of research abroad (see Fisher et al. (2010) for a complete review of techniques), the field is still incipient in Brazil, considering the text of the explanatory notes. The few existing works are more focused on lexical and volumetric analysis concerning the size of the reports, for instance, its number of words (Gomes et al., 2018), and not on their content, as proposed in this research. Still, some works have deepened the issue of legibility, which involves a quantitative analysis of the content (words, phrases, and paragraphs) to calculate the Flesch index (Gomes et al., 2018; Holtz & Santos, 2020).

Unlike a syntactic analysis, this article evaluates the semantic content of the reports, that is, the meaning of its words. We sought to explore elements related to words in their context, thus considering the purpose of the sentences and not just the quantity or frequency of the words. Therefore, one can say that there is an approximation of an analyst's reading experience and the results of the technique used. When considering the words in their context, we relate texts with similar meanings (even though they use different words), which differs from previous works that analyzed explanatory notes. The mechanism that allows this approach is the matrix of terms and documents weighted by the TF-IDF index, detailed in the next section.

Furthermore, although the area that works with the analysis of annual reports released by companies and its relationship with risk prevention and forecasting is mature, few studies bring this approach considering specifically the explanatory notes. This article fills these research gaps, proposing an analysis that considers the semantic content of explanatory notes for financial institutions. The objective is to contribute and expand previous studies that have analyzed the explanatory notes but considering a different approach. To achieve this, we used the method described in the next section.

3 Method

This research is quantitative and exploratory in nature. The objective is to analyze the effectiveness of the risk management disclosure of financial institutions, considering the content of the explanatory notes. To achieve this, we collected excerpts from the explanatory notes that are about the risk management structure, usually located at the end of the statements under the title “Risk Management” or “Management Processes.” The chosen samples were Credit, Financing, and Investment Companies (“Finance Companies”) and Development Agencies, classified in segment 4 (S4), authorized to operate by the Central Bank of Brazil (BCB). We collected the explanatory notes accompanying the financial statements were collected from the Financial Statements Center of the Financial System (*Central de Demonstrações Financeiras do Sistema Financeiro - CDSFN*), created by the Circular 3,964/2019, available on the BCB website (https://bcb.gov.br/estabilidadefinanceira/cdsfn), from June to December 2019. For older periods, we used the websites of financial institutions, the government gazette of the States and Google. All the demonstrations were in Portuguese.

In selecting our sample, we had two reasons in mind. The first was the obligation of the sampled sectors to comply with risk management regulations despite their smaller coverage area considering other financial institutions, such as commercial banks. Given the exploratory nature of this research, we chose to evaluate the proposal effectiveness from companies that would have fewer elements of complexity required by the regulations. The second reason was the fact that the two groups sampled are concise. The final sample comprises 18 Finance Companies and 14 Development Agencies, which contains almost all (86%) institutions of these sectors regulated by the BCB, allowing a comprehensive scope for the tests performed.

We captured data from the statements manually since the documents did not have a standard for automated extraction. We had to exclude five out of the 37 institutions from the sample because their data were impossible to capture: the statements were either protected or in image format with no possibility of conversion. The final period of the collection comprised statements from December 2014 to December 2019, with two annual statements, one in June and one in December of each year.

The normative taken as the basis for risk measurement was Resolution 4,557, from February 23, 2017, which refers to the mandatory implementation of risk and capital management structures for segments 1 to 4 (S1 to S4), in compliance with Resolution 4,553/17, considering the relevance and proportionality of the supervised entities. The regulation revoked Resolutions 3,380/06, 3,464/07, 3,721/09, 4,090/12, and 3,988/11, which gave specific treatment for the management of operational, market, credit and liquidity risks, and management of capital, respectively, thus consolidating them into a single regulation, introducing the concept of integrated risk management.

As the proposal of this article is innovative and does not have a comparison, quantitative risk measures were used for evaluation and benchmarking. The risk measures collected for comparison were the Basel index (ratio between the reference equity and risk-weighted assets) and the Inad 90 index (percentage of operations in the institution’s credit portfolio overdue for more than 90 days, classified at risk levels from D to H, a proxy for nonperforming assets), retrieved from the BCB website (https://www3.bcb.gov.br/ifdata/). The purpose of the comparison was to obtain a quantitative and objective view of the risk management policy of each institution since these indices express the risk appetite.

Some of the 32 sampled institutions did not have the indices in one or more of the 11 periods of analysis. For some others, there was no text about risk in the explanatory note. Either the statement was unavailable, or there was no mention of it. Hence, the final sample had 241 records (text excerpts), 133 from Finance Companies, and 108 from Development Agencies.

To analyze the content on risk in the explanatory notes, we evaluated the distance between the excerpts of risk management in the institutions and the excerpt on the risk management structure of the referred guide (Chapter I, Section III). Distance measures in texts seek to highlight the semantic proximity between different documents. For instance, we use distance measures for identifying plagiarism and analyzing consumer similarity (Hourrane & Benlahmar, 2017; Saumya et al., 2018). In this study, the distance measures seek to show how adherent the institutions are to the normative, considering that the closer they are, the better they meet aspects of the norm, and the more distant, the less their report contemplates these aspects.

The distance measure chosen was the cosine similarity. This measure is widely adopted in the text mining community and is just the cosine of the angles between two texts, projected in the same vector space (Aggarwal & Zhai, 2012; Li & Han, 2013). Its range is [0.1], and it indicates that the similarity grows the closer it is to 1. The intent is that texts whose semantic meanings are similar will be projected in the same direction, even though the size of the texts, and consequently, the size of the vectors, is different. This particularity of the cosine similarity makes its application suitable for this study since there is variability in the number of texts between the financial institutions and the analyzed guide, which were not normalized to avoid loss of information. The purpose of this analysis was to understand the ability of the explanatory note excerpt to illustrate risk management, considering that variability between the text excerpt and the resolution would demonstrate variability in risk measures.

We calculated the cosine similarity between each institution-normative pair (documents *U* and *V*) as follows (Aggarwal & Zhai, 2012):

$$cosine(U,V) = \frac{U.V}{\|U\|.\|V\|} = \frac{\sum_{i=1}^{k} U_i V_i}{\sqrt{\sum_{i=1}^{k} U_i^2}\sqrt{\sum_{i=1}^{k} V_i^2}}$$

Being *U* and *V* composed by the normalized frequency of their *i* terms. The texts were converted into normalized vectors through R text mining tools using RStudio. The first step was to assemble the 242 pieces of texts (241 excerpts of explanatory notes and one excerpt of the guide) into a *corpus* and start the pre-processing process. This pre-processing aims to reduce the dimensionality of the data and the noise that occurs when working with text data, since according to the Zipf Law, a small part of the vocabulary is enough to express a large portion of the information (Manning et al., 2008). Pre-processing included the removal of "stop words" (frequent words, but with low semantic meaning, such as articles and prepositions), the transformation to lowercase letters, and the removal of additional blank spaces, according to best practices reported in the literature (Aggarwal & Zhai, 2012).

The second step was to generate a matrix of terms and documents with lines corresponding to the unique expressions in the texts and columns corresponding to each text in the *corpus*. We decided to generate two such matrices, one for Finance Companies with dimension 2,272 x 134, and one for Development Agencies, with dimension 2,365 x 109, to respect the particularities of each group. The last column of both matrices corresponds to the resolution section.

We used the TF-IDF index (Term-Frequency and Inverse Document-Frequency) to fill the cells of each matrix. This weighing scheme values exclusive expressions and devalues frequent terms, which do not distinguish the documents because they are present in many of them, allowing for a more sophisticated analysis of the semantic space. It also improves the performance of cosine similarity (Li & Han, 2013).

We calculate the TF-IDF index as follows (Aggarwal & Zhai, 2012):

$$TF - IDF = \frac{f_{t,d}}{\max_t f_{t,d}} \times \ln\frac{N}{n_t}.$$

The first part of the formula is the relative frequency of each term, with $f_{t,d}$ being the frequency of each term t in each document d and $\max_t f_{t,d}$ being the maximum frequency of a single term in all documents. The second part of the formula is the natural logarithm of the ratio between the total number of documents (N) and the number of documents containing the given term (n_t). The objective is to balance the frequency of each term (first part) with its ability to discriminate groups of documents (second part). Frequent terms among all documents receive lower index values. Recurrent terms in fewer documents tend to receive higher values.

The vectors U considered for calculating the cosine similarity for the Finance Companies consisted of columns 1 to 133 of the respective matrix of terms and documents and the vector V is column 134 of the same matrix, representing the resolution portion. For Development Agencies, a similar process was followed, with columns 1 to 108 being the vector U and column 109 being the vector V. In this way, the vectors were of the same size and considered the presence or absence of the most relevant words, given the TF-IDF weighting used. We calculated each similarity measure individually, that is, columns $1 \times 134, 2 \times 134, \ldots, 133 \times 134$ for Finance Companies and columns $1 \times 109, 2 \times 109, \ldots, 108 \times 109$ for Development Agencies. The following section presents the analysis of the results.

4 Analysis of Results

The analyzed sample had 241 texts taken from the explanatory notes section of the statements from 32 institutions authorized to operate by the Central Bank of Brazil (18 Finance Companies and 14 Development Agencies). Although the nature of the activity of each of these classes is different, as the Finance Companies operate more focused on credit and the Development Agencies have a more social financial proposal, it is interesting to observe the data together for the objectives of this work. Thus, we divided this section into two subsections. The first subsection presents a more descriptive approach given the content of the texts, and the second one shows an analysis of the relationship between texts and risk measures.

4.1 Descriptive Analysis

The Credit, Financing and Investment Societies (SCFI), also known and here called Finance Companies, grant loans and financing the acquisition of goods, services, and working capital. They abide by the Brazilian Ministry of Economy Ordinance 309/59 and by the Central Bank of Brazil Resolution 45/66 (BCB, 59,66). They Those not linked to banks often operate as the financial arm of commercial or industrial groups, such as department stores and vehicle manufacturers. Finance Companies usually operate in niches of high risks, such as financing used vehicles and agreements with commercial establishments. Usually,

banks are not interested in such areas. They only operate by attracting funds through the issuance of their securities or bills of exchange (Viera et al., 2012). The 134 texts of the Finance Companies group were gathered in a *corpus* (set of documents) with 2,272 unique terms. The most frequent terms identify the characteristics of the excerpts from the explanatory notes as shown in Figure 1.

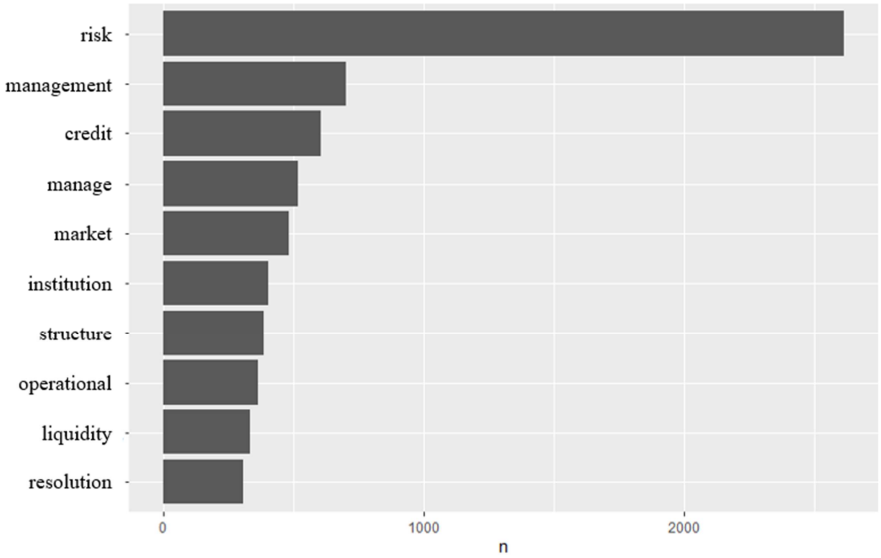

Figure 1 - Top 10 most frequent words in the Finance Companies *corpus*
Source: Research data

Although Resolution 4,557/17 (BCB, 2017) mentions six different types of risks (credit, market, interest rates, operational, liquidity, and socio-environmental variation) and the importance of including other risks relevant to each institution, only credit, market, operational, and liquidity risks are among the most frequent words. To analyze this question, we performed an analysis of *n*-grams, with *n* ranging from two to five. This analysis allows us to visualize the most frequent terms in the sentences, complementing the "bag of words" type of analyses, which do not consider the order of words. Table 1 presents the most frequent bi and pentagrams for Finance Companies.

Table 1:
Set of n-terms (n=2 and n=5) in most frequent order for Finance Companies

bigram	Freq.	pentagram	Freq.
Risk Management	566	(...) compatible the size complexity products (...)	26
Market risk	332	(...) deficiency or inadequacy internal processes (...)	26
Management Risk	330	(...) the size complexity products services (...)	26
Credit Risk	329	(...) or inadequacy internal processes people (...)	26
Operational Risk	300	(...) size complexity products services offered (...)	26
Management Framework	215	(...) failure deficiency or inadequacy processes (...)	23
Liquidity Risk	177	(...) defined as possibility occurrence losses (...)	22
CMN Resolution	109	(...) identify evaluate monitor control mitigate (...)	22
Central Bank	98	(...) risk market operational liquidity credit (...)	22
Internal Control	93	(...) liquidity risk the liquidity risk (...)	21

Source: Research data

Through Table 1, it is possible to verify that the risks of interest rate and socio-environmental variation are not present in the explanatory notes of the Finance Companies. In addition, the pentagrams recurringly mention the complexity of products and the inadequacy of processes, which may indicate the main risks not mentioned in the resolution, but relevant to this set of institutions.

The Development Agencies are unique State institutions, which operate in the financing of capital associated with Federal projects, planned in State programs for economic and social development, regulated by Resolution 2,828/01 of the Central Bank of Brazil (BCB, 2001). According to Ribeiro (2004), these agencies have restrictive legal and economic factors, and their ability to promote new projects and investments is the key to their success. The 109 texts of this group formed a corpus with 2,366 unique terms. Figure 2 shows the most frequent terms.

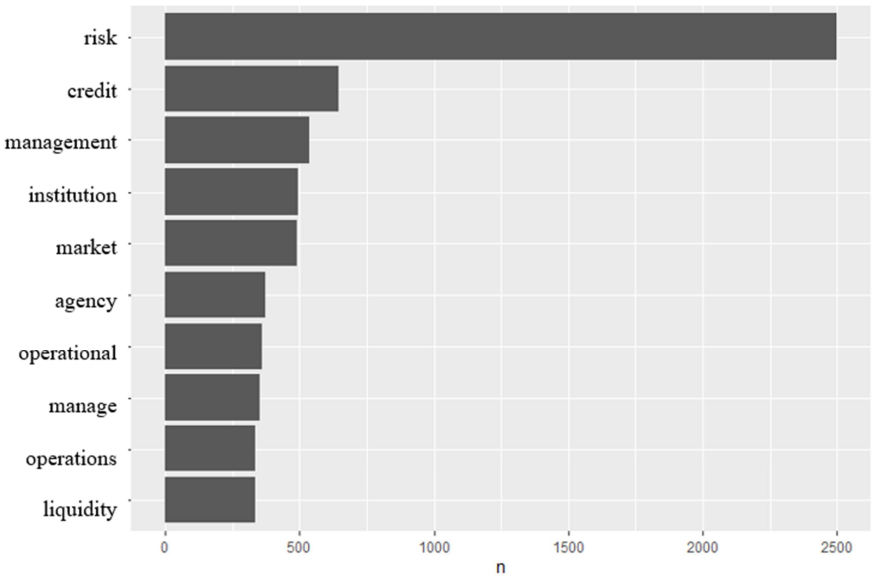

Figure 2 - Top 10 most frequent words in the Development Agencies *corpus*
Source: Research data

It is possible to notice the high similarity between the two groups concerning the most frequent terms: of the top 10 expressions, eight are the same, although the operation area of the two groups is, in principle, different. The word "agency" stands out, an expected term, given the nomination of the institutions that work in development. Equally, there is also a lack of mentions of the socio-environmental risk and the risk of interest rate variation, mentioned in Resolution 4,557/17 (BCB, 2017). The n-gram analysis results, which deepen the perception of the most frequent terms, are shown in Table 2.

Table 2:
Set of n-terms (n=2 and n=5) in most frequent order for Development Agencies

bigram	Freq.	pentagram	Freq.
Risk Management	429	(...) deficiency or inadequacy internal processes (...)	50
Credit Risk	391	(...) failure deficiency or inadequacy processes (...)	50
Operational risk	315	(...) financial obligations under agreed terms (...)	49
Market risk	266	(...) or inadequacy internal processes people (...)	49
Management Risk	208	(...) associated to borrower compliance or (...)	47
Money Laundering	190	(...) internal people systems or events (...)	47
Liquidity Risk	175	(...) losses associated to borrower compliance (...)	47
Losses Occurrence	132	(...) people systems or external events (...)	47
Possibility Occurrence	121	(...) possibility occurrence losses resulting failure (...)	47
Resolution n	113	(...) occurrence losses resulting failure deficiency (...)	46
Central Bank	105	(...) crime prevention of laundering money (...)	39

Source: Research data

The highest variability of terms found in Table 2 regarding Table 1 shows that this group (Development Agencies) presents a broader discussion in the explanatory notes concerning the risk management structure. Although it may reflect the different actions of the institutions in this group, it is clear that money laundering is a concern. Still, there is a similarity in the inadequacy (or deficiency, a more present term in this group) of processes.

4.2 Risk *versus* Text Variability

In this subsection, Finance Companies and Development Agencies explanatory notes are evaluated as a source of data to understand the position of these institutions on their risk policies. The assumption is that the explanatory notes present reliable information that provides greater transparency about the organizations' positions regarding risk management since the explanatory notes must be a narrative description of the policies adopted by the companies (CPC, 2020; IFRS, 2020). We expect that any policy change, regardless of its motivators being internal or external, would entail substantial changes in the explanatory notes. That is particularly true if the risk appetite increases or decreases: we should expect changes in the explanatory notes. After all, the explanatory notes need to provide material and relevant information to the public (CPC, 2014).

To understand the relationship between texts and risk behavior, we first verified whether text similarity and risk measures varied jointly. Figure 3 presents the Basel index, the Inad 90, and the cosine similarity between the Finance Companies (F1 to F18) text excerpts and the resolution. The graph plots all indexes on a relative scale (from 0 to 1). One can observe that the indexes do not seem connected to the

content of the texts about the resolution, except for a few institutions (especially F3, F4, and F5) in which one can see a pattern.

We can conclude the same from Figure 4, which presents the corresponding analysis for Development Agencies. Without exceptions, the risk measures are aligned in terms of variability but detached from the cosine similarity movement, which seems to oscillate very little compared to the risk itself. This low relationship raises the hypothesis that the texts do not vary as much as the risk indexes themselves.

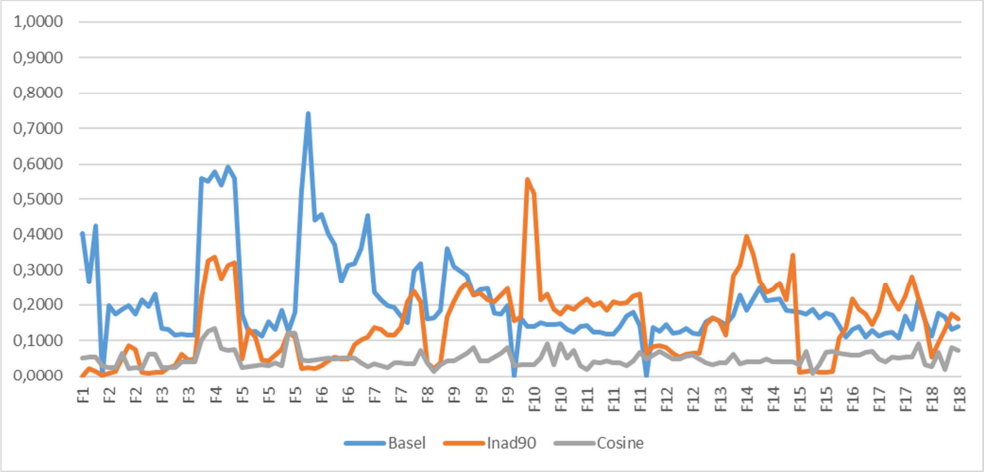

Figure 3 - Basel Index, Inad 90 and Cosine Similarity for Finance Companies
Source: Research data

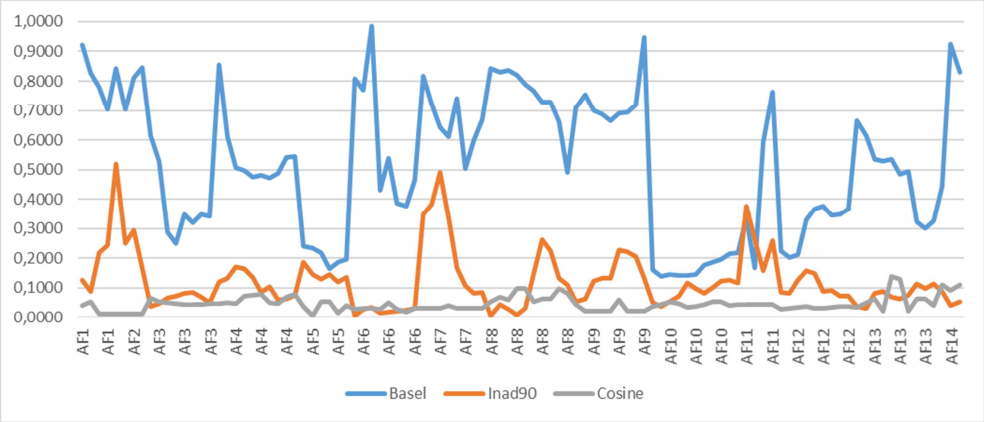

Figure 4 - Basel Index, Inad 90 and Cosine Similarity for Development Agencies
Source: Research data

To deepen our analysis of this hypothesis, we examined the dispersion between the similarities. Considering that the texts should inform about risk policies and knowing that this risk varies over time, we expected that there would be an intra-organizational dispersion. That would appear as a difference between one company's text excerpts and the text excerpt of the resolution over time. Figures 5 presents the cosine similarities of all 133 Finance Companies text excerpts and the resolution text. Similarly, Figure 6 depicts the statistics for all 108 Development Agencies text excerpts.

Figures 5 and 6 show a pattern in the graphs that does not confirm the expected output for nearly all institutions. There is generally low dispersion between the cosine dissimilarities of different text excerpts and the resolution for almost all firms, which might denote a more generic text used to inform management and risk policies. We highlighted in the Figures text excerpts of the same institution with very low variability, which indicates a highly similar text. We must remember that the standard text that refers to risk management used to measure similarity is the same in all comparisons. These two shreds of evidence reinforce the hypothesis that the explanatory notes do not represent the reality of risk management policy, given the evidence of indices variability.

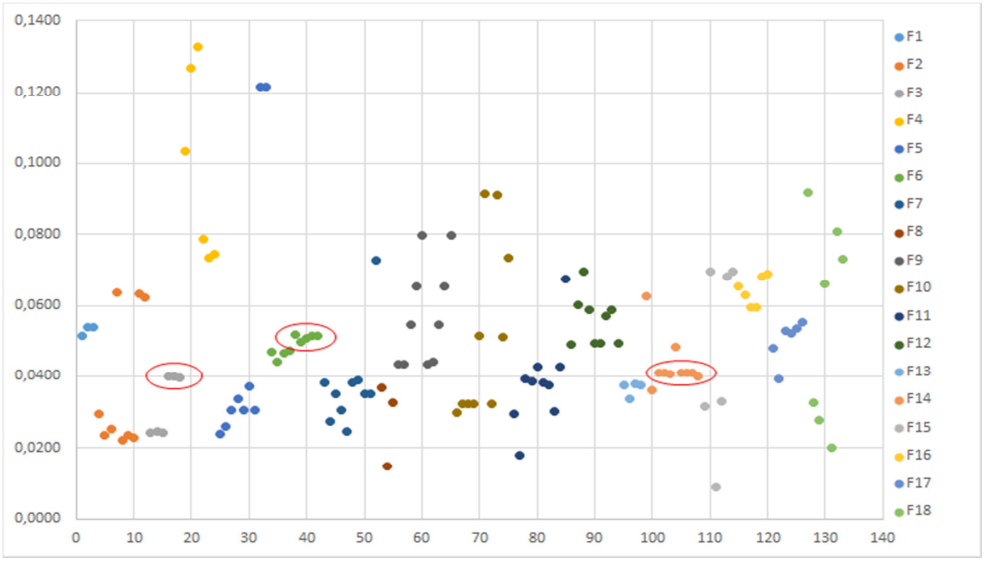

Figure 5 - Cosine similarity values in Finance Companies
(colors represent the 18 analyzed companies, highlighted are the institutions with very low variability)
Source: Research data

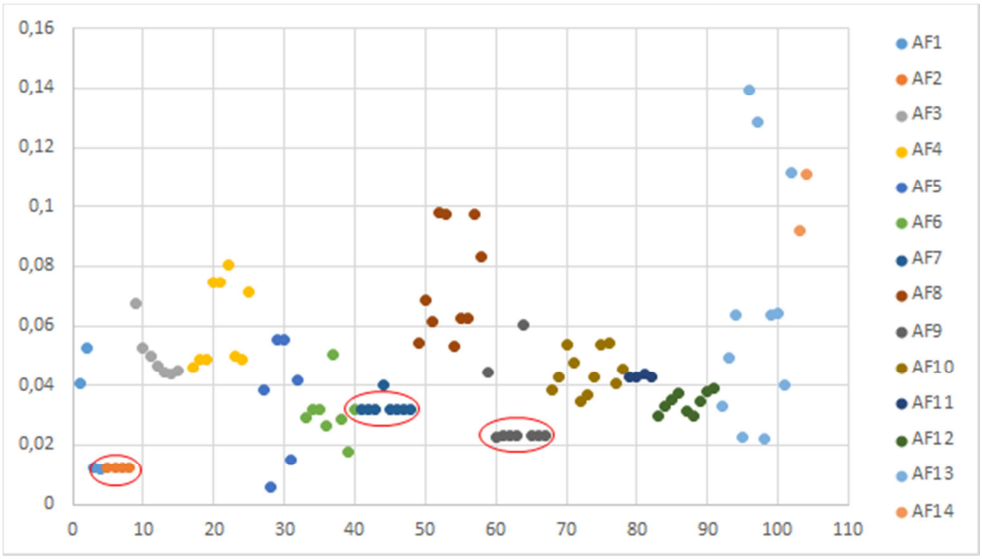

Figure 6 - Cosine similarity values in Development Agencies
(colors represent the 14 analyzed agencies, highlighted are institutions with very low variability)
Source: Research data

To expand the analysis of the relationship between the structure of the explanatory notes text excerpts and the risk measures, we present Figures 7 and 8. We selected two institutions with the highest number of records (text excerpts and risk measures) in each group, one with the highest total variability (the longest distance between the passages of text) and one with the smallest total variability, to avoid bias.

The objective was to understand how the variability of risk measures was (or was not) associated with the variability between texts from the same institution. We expected that changes between the risk measure periods would be associated with greater distances between the text excerpts in the respective statements since the explanatory notes are the space to better describe this type of change (IFRS, 2020). However, we could not observe such a relationship: the figures show that although there is variability in risk measures between one period and the next, the text has remained very similar.

In Figure 7, it is possible to notice that both institutions have variability in risk measures, but they do not have support in text variability. Institution F14 showed variability in its excerpts between Dec/14 and Dec/16. However, from Dec/16 to Jun/19, the text was very alike, with virtually zero variability. Institution F10, on the other hand, presented remarkably similar texts between Jan/14 and Jun/17 even with the high variability in the Inad 90 risk measure from the same period.

In Figure 8, a similar phenomenon occurs. It is noteworthy that the relative balance of risk measures contrasts with the oscillation of the texts, which is not in itself an indicator of loss of information since the firm might change its risk management policy to stabilize its own risk. However, of risk variability, as seen in AF10 between Dec/16 and Dec/17, are not reflected in the text.

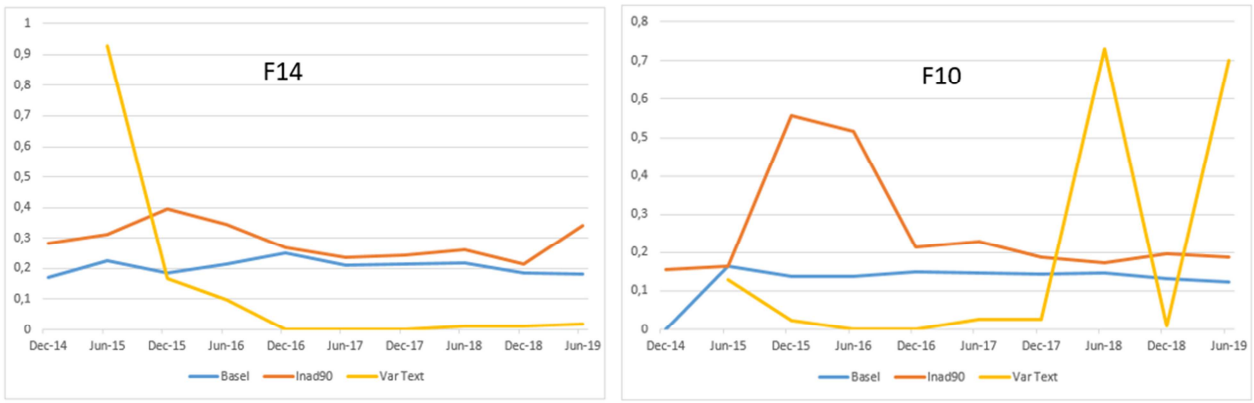

Figure 7 - Variability of Risk Measures *versus* Text from "Finance Companies" F14 and F10
(The yellow line represents the text variability, that is, $1 - cosine(U,V)$)
Source: Research data

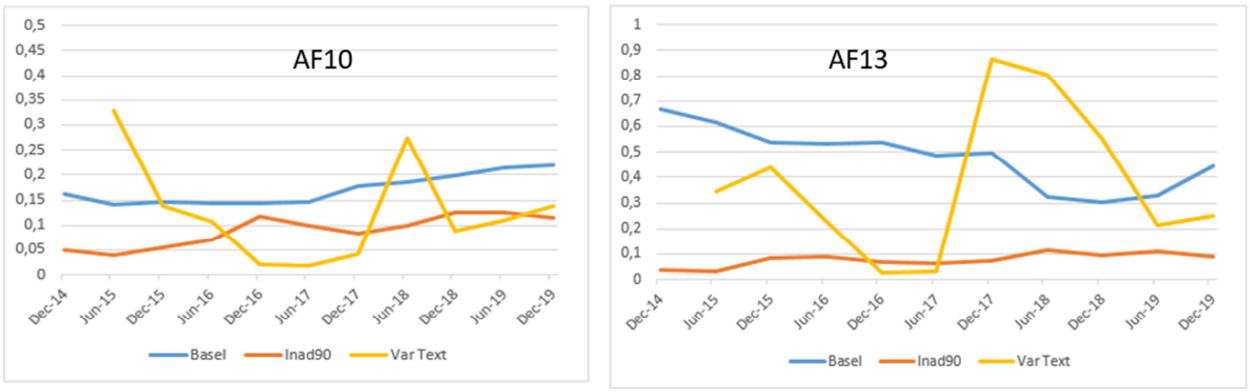

Figure 8 - Variability of Risk Measures *versus* Text from Development Agencies AF10 and AF13
(The yellow line represents the text variability, that is, $1 - cosine(U,V)$)
Source: Research data

Considering the difference in the observed variabilities, we decided to delve into some individual cases. Thus, the same institutions in Figures 7 and 8 were analyzed using a heat map (Figures 9 and 10). The heat map shows the similarities of the text excerpts between the periods. As numbers closer to 1 present strong signals of text reuse, these were signaled with warm colors, while smaller values (passages more distant from each other) are in the green scale.

	Dec-14	Jun-15	Dec-15	Jun-16	Dec-16	Jun-17	Dec-17	Jun-18	Dec-18	Jun-19
Dec-14		0,07	0,07	0,07	0,07	0,07	0,07	0,06	0,07	0,07
Jun-15	0,07		0,83	0,75	0,75	0,75	0,75	0,74	0,75	0,74
Dec-15	0,07	0,83		0,90	0,90	0,90	0,90	0,89	0,89	0,88
Jun-16	0,07	0,75	0,90		1,00	1,00	1,00	0,99	1,00	0,98
Dec-16	0,07	0,75	0,90	1,00		1,00	1,00	0,99	1,00	0,98
Jun-17	0,07	0,75	0,90	1,00	1,00		1,00	0,99	1,00	0,98
Dec-17	0,07	0,75	0,90	1,00	1,00	1,00		0,99	1,00	0,98
Jun-18	0,06	0,74	0,89	0,99	0,99	0,99	0,99		0,99	0,97
Dec-18	0,07	0,75	0,89	1,00	1,00	1,00	1,00	0,99		0,98
Jun-19	0,07	0,74	0,88	0,98	0,98	0,98	0,98	0,97	0,98	

F14

	Dec-14	Jun-15	Dec-15	Jun-16	Dec-16	Jun-17	Dec-17	Jun-18	Dec-18	Jun-19
Dec-14		0,87	0,89	0,89	0,89	0,87	0,89	0,26	0,26	0,78
Jun-15	0,87		0,98	0,98	0,98	1,00	0,98	0,27	0,27	0,89
Dec-15	0,89	0,98		1,00	1,00	0,97	1,00	0,27	0,27	0,87
Jun-16	0,89	0,98	1,00		1,00	0,97	1,00	0,27	0,27	0,87
Dec-16	0,89	0,98	1,00	1,00		0,97	1,00	0,27	0,27	0,87
Jun-17	0,87	1,00	0,97	0,97	0,97		0,97	0,27	0,27	0,89
Dec-17	0,89	0,98	1,00	1,00	1,00	0,97		0,27	0,27	0,87
Jun-18	0,26	0,27	0,27	0,27	0,27	0,27	0,27		0,99	0,30
Dec-18	0,26	0,27	0,27	0,27	0,27	0,27	0,27	0,99		0,30
Jun-19	0,78	0,89	0,87	0,87	0,87	0,89	0,87	0,30	0,30	

F10

Figure 9 - Similarity between the texts of the Finance Companies F14 (above) and F10 (below)
(The diagonal was removed for better viewing, and it contained the value of 1, which indicates maximum similarity)
Source: Research data

Figure 9 shows the opposite trend of the two institutions. While F10 seems to be modifying its texts more recently, F14 seems to have its writings increasingly similar. Both, however, have values of 1.00 outside the diagonal, which indicates the use of the same text excerpts. Although it is reasonable to consider that the risk management policy will not change every six months, the similar texts do not show variability. Therefore, there is no relevant information for the decision-maker, which contradicts the role of this type of data in financial statements (Fisher et al., 2010).

	Dec-14	Jun-15	Dec-15	Jun-16	Dec-16	Jun-17	Dec-17	Jun-18	Dec-18	Jun-19	Dec-19
Dec-14		0,67	0,75	0,67	0,64	0,66	0,63	0,51	0,45	0,49	0,49
Jun-15	0,67		0,86	0,78	0,78	0,78	0,76	0,68	0,71	0,66	0,70
Dec-15	0,75	0,86		0,89	0,86	0,88	0,85	0,65	0,57	0,63	0,63
Jun-16	0,67	0,78	0,89		0,98	0,99	0,96	0,75	0,66	0,73	0,62
Dec-16	0,64	0,78	0,86	0,98		0,98	0,97	0,74	0,67	0,72	0,61
Jun-17	0,66	0,78	0,88	0,99	0,98		0,96	0,76	0,67	0,74	0,62
Dec-17	0,63	0,76	0,85	0,96	0,97	0,96		0,73	0,65	0,71	0,62
Jun-18	0,51	0,68	0,65	0,75	0,74	0,76	0,73		0,91	0,98	0,88
Dec-18	0,45	0,71	0,57	0,66	0,67	0,67	0,65	0,91		0,89	0,86
Jun-19	0,49	0,66	0,63	0,73	0,72	0,74	0,71	0,98	0,89		0,86
Dec-19	0,49	0,70	0,63	0,62	0,61	0,62	0,62	0,88	0,86	0,86	

AF10

	Dec-14	Jun-15	Dec-15	Jun-16	Dec-16	Jun-17	Dec-17	Jun-18	Dec-18	Jun-19	Dec-19
Dec-14		0,65	0,87	0,69	0,69	0,69	0,02	0,66	0,21	0,19	0,22
Jun-15	0,65		0,56	0,45	0,45	0,45	0,02	0,43	0,13	0,11	0,14
Dec-15	0,87	0,56		0,77	0,74	0,72	0,16	0,61	0,22	0,20	0,23
Jun-16	0,69	0,45	0,77		0,97	0,96	0,13	0,63	0,34	0,30	0,31
Dec-16	0,69	0,45	0,74	0,97		0,97	0,13	0,64	0,35	0,30	0,32
Jun-17	0,69	0,45	0,72	0,96	0,97		0,13	0,65	0,34	0,29	0,32
Dec-17	0,02	0,02	0,16	0,13	0,13	0,13		0,20	0,07	0,25	0,07
Jun-18	0,66	0,43	0,61	0,63	0,64	0,65	0,20		0,45	0,37	0,44
Dec-18	0,21	0,13	0,22	0,34	0,35	0,34	0,07	0,45		0,79	0,87
Jun-19	0,19	0,11	0,20	0,30	0,30	0,29	0,25	0,37	0,79		0,75
Dec-19	0,22	0,14	0,23	0,31	0,32	0,32	0,07	0,44	0,87	0,75	

AF13

Figure 10 - Similarity between the texts of the Development Agencies AF10 (above) and AF13 (below)
(The diagonal was removed for better viewing, and it contained the value of 1, indicating maximum similarity)
Source: Research data

Figure 10 highlights an even higher similarity in the writings of the selected agencies. Even though AF10 texts have become more distant over time (higher presence of green indicators at the edges), the texts between the periods are very similar (concentration of red diagonally), especially between Jun/16 and Dec/17. We raise the hypothesis that periods with the same external company advising or auditing the statements may be the source of the highest similarity between closer periods. The same happens with AF13, although without the same longitudinal pattern. The pattern concentrates on some periods, such as from Dec/15 to Jun/17 and Dec/18 to Dec/19.

The findings allow us to conclude that there is a low relationship between the text and risk metrics. Figures 5 and 6 made it possible to raise the hypothesis of low text variability from different periods in each institution, confirmed by the cases analyzed in Figures 7 and 8. We notice that the notes examined do not fully comply with OCPC 07 (CPC, 2014). The characteristic of utility for the user in decision-making processes is impaired since there is a low relationship with the risk metrics. In addition, there is an excess of irrelevant information, given the similarity of the texts between periods, which can cause a deviation from the final goal of informing the user.

These results are also in line with the studies by Ryan (2015), who classified the qualitative risk information of American financial institutions as opaque and irrelevant. The Basel and Inad 90 indices show a variation in the risk policy of the institutions over the period. However, this variation cannot be evidenced in the notes, indicating a gap between the texts on risk management in the explanatory notes and the implemented risk management policy, or even some level of reuse of texts, disconnected from the events and decisions of managers during that period. To better illustrate this evidence, we selected some excerpts with high similarity to perform a counter-test of the low variability evidenced. Table 3 displays the results.

The exact repetition in the excerpts highlights the low informational value of the source that should assist decision-making processes. Besides not having the correct curation of the content on risk management, there is a detachment from the practice since the texts do not reflect the fluctuations found in the quantitative indices. The explanatory notes, which should assist in understanding the adopted policies that led to the results found in the writings, in the analyzed cases, often have a figurative role, which can harm the users of this information. Hence, we can not recognize the explanatory notes, in the analyzed cases, as a space for the expansion of accountability and transparency in the accounting reports so that they

are reliable representations of the processes used by companies regarding risk management (Hendriksen & Van Breda, 2010; Nakagawa et al., 2007).

Table 3:
Repeated text excerpts in Explanatory Notes

Code	Date	Text excerpt from Explanatory Notes
F14	Jun 16	"Operational, Market and Credit Risk Management Structure: The [F14] Economic and Financial Conglomerate, in compliance with the provisions of National Monetary Council Resolution No. 3.380/06, 3.464/07 and 3.721/09 and its amendments, has an Operational, Market and Credit risk management structure, respectively, segregated from the business and internal auditing areas, as well as being able to identify, assess, monitor, control and mitigate their risks, including those arising from outsourced services. All risk management structures are directed (...)"
F14	Dec 16	"Operational, Market and Credit Risk Management Structure: The [F14] Economic and Financial Conglomerate, in compliance with the provisions of National Monetary Council Resolution No. 3.380/06, 3.464/07 and 3.721/09 and its amendments, has an Operational, Market and Credit risk management structure, respectively, segregated from the business and internal auditing areas, as well as being able to identify, assess, monitor, control and mitigate their risks, including those arising from outsourced services. All risk management structures are directed (...)"
F14	Jun 17	"Operational, Market and Credit Risk Management Structure: The [F14] Economic and Financial Conglomerate, in compliance with the provisions of National Monetary Council Resolution No. 3.380/06, 3.464/07 and 3.721/09 and its amendments, has an Operational, Market and Credit risk management structure, respectively, segregated from the business and internal auditing areas, as well as being able to identify, assess, monitor, control and mitigate their risks, including those arising from outsourced services. All risk management structures are directed (...)"
AF10	Dec 16	"(...) Summary of the actions developed within the scope of risk management: a. Market risk: The Institution's Market Risk Management Policy is reviewed annually to identify, assess, monitor, and control the exposure of these risks, ensuring the existence of efficient mechanisms for supervision and monitoring. The market risk is not significant for [AF10], considering that most of its operations have 'matched' rates on assets and liabilities of lending operations, which mitigates the exposure risk. b. Credit risk: Credit risk arises from the possibility of losses (...)"
AF10	Jun 17	"(...) Summary of the actions developed within the scope of risk management: a. Market risk: The Institution's Market Risk Management Policy is reviewed annually to identify, assess, monitor, and control the exposure of these risks, ensuring the existence of efficient mechanisms for supervision and monitoring. The market risk is not significant for [AF10], considering that most of its operations have 'matched' rates on assets and liabilities of lending operations, which mitigates the exposure risk. b. Credit risk: Credit risk arises from the possibility of losses (...)"
AF10	Dez 17	"(...) Summary of the actions developed within the scope of risk management: a. Market risk: The Institution's Market Risk Management Policy is reviewed annually to identify, assess, monitor, and control the exposure of these risks, ensuring the existence of efficient mechanisms for supervision and monitoring. The market risk is not significant for [AF10], considering that most of its operations have 'matched' rates on assets and liabilities of lending operations, which mitigates the exposure risk. b. Credit risk Credit risk arises from the possibility of losses (...)"

Source: Research data

Given the evidence that the excerpt from the explanatory notes dedicated to clarifying the risk policy is not only not related to risk appetite indices, it also sometimes does not present any new information, and one can perceive the existence of managers' discretion in the disclosure (and omission) of information on risk management (Verrecchia, 2001). The repetition detected by the maximum similarity index brings to light a latent need for incentives or requirements towards the usefulness of this information. Because they are financial institutions, there is an increase in these concerns since ignorance of risks can be very harmful to users (Dantas et al., 2005).

5 Final Considerations

This article aimed to analyze the effectiveness of risk management communication based on the content of explanatory notes, using text mining techniques. The text of the explanatory notes has the function of presenting additional data qualitatively, which helps in the understanding and reading of the financial statements. In particular, the explanatory notes can help clarify elements such as risk appetite, management measures, and contingency structure. Thus, we tried to expand the research possibilities by considering the explanatory notes as more than just lexical and volumetric analyses and using their content.

We used the cosine similarity measure to gauge the proximity of the excerpts on risk management in the explanatory notes from two groups of financial institutions (Finance Companies and Development Agencies) with Resolution 4,557/17 of the CBC, which regulates integrated risk management. The objective was to identify if there was any relationship with traditional risk measures. The intent was that the text could evidence the risk management policy, which impacts the related quantitative indexes.

To compare the text with the risk management policies practiced by the institutions, we compared the cosine similarity with two other traditional risk metrics: the Basel index and the Inad 90 index. In general, there was no evidence to support a relationship between cosine similarity and the risk indices. Additionally, we analyzed the variability of the text itself, and the results showed that the frequency of repeated texts is high, with some excerpts being equivalent, which raises some reflections.

The text selected for the analysis deals exclusively with risks. Since it is present in the explanatory notes, we expected that it would reflect the positioning and the defense of the risk in the organizations, which is variable throughout the analyzed period. However, the low relationship of text contents with the

quantitative indices denotes a possible distance between this information and the current risk management practice in organizations. In a way, institutions may be reporting certain practices and performing others.

Based on the evidenced results, we perceived that the text is reused or slightly modified, which causes a low variability, within the same institution, between periods, differently from what occurs with the risk indices studied. In this scenario, the text about risk in the explanatory notes would function to meet a requirement without relevant information for the stakeholders. Hence, in the analyzed case, the content of the explanatory notes does not reflect their objective of reliably describing the risk management policies of these organizations, increasing transparency between the company and users of accounting and financial information.

This article presents some limitations that we sought to overcome whenever possible. The unavailability of some demonstrations and some indicators caused the base to be unbalanced, which led some institutions to have more records than others. The choice of distance measurement matters a lot, and a test with other metrics and techniques (such as Euclidean distance and pre-text vectorization) is on the horizon of future studies. Incorporating other risk indices and other financial institutions that comply with Resolution 4,557/17. Finally, the expansion of text sources (such as administration reports or press releases) can help better investigate the relationship between disclosed information and risk indices.

References

Aggarwal, C. C., & Zhai, C. (2012). A survey of text clustering algorithms. In *Mining text data* (pp. 77-128). Springer, Boston, MA. https://doi.org/10.1007/978-1-4614-3223-4_4

BCB – Banco Central do Brasil. (2012). *Resolução nº 4.144*, de 2012. 27 de setembro de 2012. Available at https://www.bcb.gov.br/pre/normativos/busca/downloadNormativo.asp?arquivo=/Lists/Normativos/Attachments/49098/Res_4144_v1_O.pdf

BCB – Banco Central do Brasil. (2017). *Resolução nº 4.557*, de 23 de fevereiro de 2017. Available at https://www.bcb.gov.br/pre/normativos/busca/downloadNormativo.asp?arquivo=/Lists/Normativos/Attachments/50344/Res_4557_v1_O.pdf

BCB – Banco Central do Brasil. (1966). *Resolução nº 45*, de 30 de dezembro de 1966. Available at https://www.bcb.gov.br/pre/normativos/busca/downloadNormativo.asp?arquivo=/Lists/Normativos/Attachments/40040/Res_0045_v8_P.pdf

BCB – Banco Central do Brasil. (1959). *Portaria do Ministério da Fazenda nº 309*, de 30 de novembro de 1959. Available at https://www.bcb.gov.br/pre/acordos_e_convenios/Portaria_MF_309-1959.pdf

BCBS - Basel Committee on Banking Supervision. (1997). Core principles for effective banking supervision. *Basileia, Suíça: Bank for International Settlements*. Available at https://www.bis.org/publ/bcbs30a.pdf

Bergamini Junior, S. (2005). Controles internos como um instrumento de governança corporativa. *Revista do BNDES*, 12(24), 149-188. Available at: http://web.bndes.gov.br/bib/jspui/handle/1408/13474.

Brown, N. C., Crowley, R. M., & Elliott, W. B. (2020). What are you saying? Using topic to detect financial misreporting. *Journal of Accounting Research*, 58(1), 237-291. https://doi.org/10.1111/1475-679X.12294

Committee of Sponsoring Organizations of the Treadway Commission - COSO (2004). *Enterprise Risk Management - Integrated Framework*. AICPA: New York, USA. Available at https://www.coso.org/documents/Framework%20Reference%20Secured.pdf

CPC – Comitê de Pronunciamentos Contábeis. *CPC 26 (R1)*. (2011) Available at: http://static.cpc.aatb.com.br/Documentos/312_CPC_26_R1_rev%2014.pdf. Accessed July 23rd, 2020.

CPC – Comitê de Pronunciamentos Contábeis. OCPC 07. (2014) Available at: http://static.cpc.aatb.com.br/Documentos/488_OCPC_07_Orienta%C3%A7%C3%A3o.pdf. Accessed March 25th, 2021.

Crockford, G. N. (1982). The bibliography and history of risk management: some preliminary observations. *Geneva Papers on Risk and Insurance*, 169-179. Available at https://www.jstor.org/stable/41950036

Dantas, J. A., Zendersky, H. C., Santos, S. C. & Niyama, J. K. (2005). A dualidade entre os benefícios do disclosure e a relutância das organizações em aumentar o grau de evidenciação. *Revista Economia & Gestão*, 5(11), 56-76. Available at http://periodicos.pucminas.br/index.php/economiaegestao/article/view/40/34

Del Gaudio, B. L., Megaravalli, A. V., Sampagnaro, G., & Verdoliva, V. (2020). Mandatory disclosure tone and bank risk-taking: Evidence from Europe. *Economics Letters*, *186*, 108531. https://doi.org/10.1016/j.econlet.2019.108531

Ding, K., Lev, B., Peng, X., Sun, T. & Vasarhelyi, M. (2018). On Creating Accounting Estimates using Machine Learning. *SSRN Electronic Journal*.

Dyer, T.; Lang, M.; & Stice-Lawrence, L. (2017) The Evolution of 10-K Textual Disclosure: Evidence from Latent Dirichlet Allocation. *Journal of Accounting and Economics* 64, 221–45. https://doi.org/10.1016/j.jacceco.2017.07.002

Elamer, A. A., Ntim, C. G., Abdou, H. A., Owusu, A., Elmagrhi, M., & Ibrahim, A. E. A. (2021). Are bank risk disclosures informative? Evidence from debt markets. *International Journal of Finance & Economics*, *26*(1), 1270-1298. https://doi.org/10.1002/ijfe.1849

Fisher, I. E., Garnsey, M. R., & Hughes, M. E. (2016). Natural language processing in accounting, auditing and finance: A synthesis of the literature with a roadmap for future research. *Intelligent Systems in Accounting, Finance and Management*, 23(3), 157-214. https://doi.org/10.1002/isaf.1386

Fisher, I. E., Garnsey, M. R., Goel, S., & Tam, K. (2010). The role of text analytics and information retrieval in the accounting domain. *Journal of Emerging Technologies in Accounting*, 7(1), 1-24. https://doi.org/10.2308/jeta.2010.7.1.1

Gomes, M., Ferreira, R. R., & Martins, V. A. (2018). O Impacto da OCPC 07 sobre o Tamanho e a Legibilidade das Notas Explicativas de Companhias Brasileiras. *Revista Universo Contábil*, 14(2), 162-184. doi: https://doi.org/10.4270/ruc.2018216

Hendriksen, E. S., & Van Breda, M. F. (2010) *Teoria da Contabilidade*. 5. ed. São Paulo: Editora Atlas.

Holtz, L., & dos Santos, O. M. (2020). Legibilidade das notas explicativas das empresas brasileiras de capital aberto. *Enfoque: Reflexão Contábil*, 39(1), 57-73. https://doi.org/10.4025/enfoque.v39i1.45275

Hourrane, O., & Benlahmar, E. H. (2017). Survey of plagiarism detection approaches and big data techniques related to plagiarism candidate retrieval. In *Proceedings of the 2nd international Conference on Big Data, Cloud and Applications* (pp. 1-6). https://doi.org/10.1145/3090354.3090369

Huang, A.; Lehavy, R., Zang, A. & Zheng, R. (2018). Analyst Information Discovery and Information Interpretation Roles: A Topic Modeling Approach. *Management Science,* 64(6), 2473–2972. https://doi.org/10.1287/mnsc.2017.2751

IFRS. *IAS1*. Available at: https://www.ifrs.org/issued-standards/list-of-standards/ias-1-presentation-of-financial-statements/. Accessed July 23rd, 2020.

Jizi, M. I., & Dixon, R. (2017). Are risk management disclosures informative or tautological? Evidence from the US banking sector. *Accounting Perspectives*, *16*(1), 7-30. https://doi.org/10.1111/1911-3838.12134

Kim, J., Kim, M., & Kim, Y. (2019). Bank transparency and the market's perception of bank risk. *Journal of Financial Services Research*, 1-28. https://doi.org/10.1007/s10693-019-00323-7

Li, B., & Han, L. (2013). Distance weighted cosine similarity measure for text classification. *In International Conference on Intelligent Data Engineering and Automated Learning* (pp. 611-618). Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-41278-3_74

Loughran, T., & McDonald, B. (2016). Textual analysis in accounting and finance: A survey. *Journal of Accounting Research*, 54(4), 1187-1230. https://doi.org/10.1111/1475-679X.12123

McAfee, A., Brynjolfsson, E., Davenport, T. H., Patil, D. J., & Barton, D. (2012). Big data: the management revolution. *Harvard business review*, 90(10), 60-68. Available at https://wiki.uib.no/info310/images/4/4c/McAfeeBrynjolfsson2012-BigData-TheManagementRevolution-HBR.pdf

Manning, C. D., Schütze, H., & Raghavan, P. (2008). *Introduction to information retrieval.* Cambridge university press.

Nahar, S., Azim, M. I., & Hossain, M. M. (2020). Risk disclosure and risk governance characteristics: evidence from a developing economy. *International Journal of Accounting & Information Management, 28*(4). 577-605. https://doi.org/10.1108/IJAIM-07-2019-0083

Nakagawa, M., Sordi Relvas, T. R., & Dias Filho, J. (2008). Accountability: a razão de ser da Contabilidade. *Revista de Educação e Pesquisa em Contabilidade (REPeC)*, *1*(3), 83-100. http://dx.doi.org/10.1590/S1519-70772007000200001

Nielsen, S., & Fuertes-Olivera, P. A. (2013). Development in lexicography: From polyfunctional to monofunctional accounting dictionaries. *Lexikos*, 23, 323-347. https://doi.org/10.5788/23-1-1218

Oino, I. (2019). Do disclosure and transparency affect bank's financial performance?. *Corporate Governance: The International Journal of Business in Society*. https://doi.org/10.1108/CG-12-2018-0378

Potter, B. N. (2005). Accounting as a social and institutional practice: Perspectives to enrich our understanding of accounting change. *Abacus*, 41(3), 265-289. https://doi.org/10.1111/j.1467-6281.2005.00182.x

Raghavan, R. S. (2003). Risk management in banks. *Chartered Accountant-New Delhi-*, *51*(8), 841-851. Available at http://un.uobasrah.edu.iq/lectures/1784.pdf

Rahrovi, D. A., Foroghi, D., & Kiani, G. H. (2019). Detecting manager's fraud risk using text analysis: evidence from Iran. *Journal of Applied Accounting Research*, 20(2), 154-171. https://doi.org/10.1108/JAAR-01-2018-0016

Ryan, S. G. (2012). Risk reporting quality: Implications of academic research for financial reporting policy. *Accounting and business research*, 42(3), 295-324. https://doi.org/10.1080/00014788.2012.681855

Ribeiro, P. A. N. (2004). Agências de fomento – ambientes institucional, legal e operacional. *Revista da Desenbahia*, Salvador, 1. Available at https://www.desenbahia.ba.gov.br/publitao/arquivos/arquivos/171dcdd029e944b1b144791bd7b91e32.pdf

Santos, N. N. F., Flores, E., Guimarães, I. P., Slomski, V. G., & Silva, A. F. (2010). Nível de evidenciação obtido na divulgação das notas explicativas de instrumentos financeiros derivativos segundo a deliberação CVM n 550/08. *Revista Contemporânea de Contabilidade*, *7*(13), 175-196. https://doi.org/10.5007/2175-8069.2010v7n13p175

Saumya, S., Singh, J. P., Baabdullah, A. M., Rana, N. P., & Dwivedi, Y. K. (2018). Ranking online consumer reviews. *Electronic Commerce Research and Applications*, 29, 78-89. https://doi.org/10.1016/j.elerap.2018.03.008

Sawanobori, C. (2013). Influence of fundraising infeasibility on accounting changes an investigation into early financial reporting practices using text mining. *Journal of Accounting and Organizational Change*, 9(4), 471-489. https://doi.org/10.1108/JAOC-08-2010-0038

Schmitt, E. (2017). Effective risk culture in banks: Responsibilities and boundaries of the risk management. *European Financial Systems 2017 - Proceedings of the 14 th International Scientific Conference*, 239-246. Available at https://is.muni.cz/do/econ/sborniky/70896034/EFS2017-Proceedings_2_final.pdf

Swandi, I., Suwardi, I. S., & Maulidevi, N. U. (2017). Automatic data interpretation in accounting information systems based on ontology. *Telkomnika (Telecommunication Computing Electronics and Control), 15*(4), 1817-1829. https://doi.org/10.12928/TELKOMNIKA.v15i4.6414

Torres, M. C., & Galdi, F. C. (2013). Análise da aderência do disclosure da gestão de riscos pelas instituições financeiras à circular Bacen 3.477. *Revista Contemporânea de Contabilidade*, *10*(21), 137-158. https://doi.org/10.5007/2175-8069.2013v10n21p137

Verrecchia, R. E. (2001). Essays on disclosure. *Journal of Accounting and Economics*, 32(1-3), 97-180. https://doi.org/10.1016/S0165-4101(01)00025-8

Yamamoto, M. M., & Salotti, B. M. (2006). *Informação contábil: estudos sobre a sua divulgação no mercado de capitais*. São Paulo: Editora Atlas.

Yang, M., Adomavicius, G., Burtch, G., & Ren, Y. (2018). Mind the gap: Accounting for measurement error and misclassification in variables generated via data mining. *Information Systems Research*, 29(1), 4-24. https://doi.org/10.1287/isre.2017.0727

NOTES

ACKNOWLEDGMENT
Does not apply.

AUTHORSHIP CONTRIBUTION
Conception and elaboration of the manuscript: C.B. Marcolin, E. D. Riva, F. S. Momo.
Data collection: E. D. Riva
Data analysis: C.B. Marcolin, J. L. Becker.
Discussion of the results: C.B. Marcolin, E. D. Riva, F. S. Momo, J. L. Becker.
Review and approval: C.B. Marcolin, E. D. Riva, F. S. Momo, J. L. Becker

DATASET
The dataset that supports the results of this study is not publicly available.

FINANCING
This work was carried out with the support of the Coordination for the Improvement of Higher Education Personnel - Brazil (CAPES) - and the National Council for Scientific and Technological Development (CNPq). In accordance with Ordinance No. 206 of September 4, 2018, "this work was carried out with the support of the Coordination for the Improvement of Higher Education Personnel - Brazil (CAPES) - Financing Code 001".

CONSENT TO USE IMAGE
Does not apply.

APPROVAL OF THE RESEARCH ETHICS COMMITTEE
Does not apply.

CONFLICT OF INTERESTS
Does not apply.

PUBLISHER
Federal University of Santa Catarina. Accounting Sciences Course and Postgraduate Program in Accounting. Publication on the UFSC Journal Portal. The ideas expressed in this article are the responsibility of their authors, and do not necessarily represent the opinion of the editors or the university.

EDITORS
Carlos Eduardo Facin Lavarda and Suliani Rover

HISTORIC
Received on: 05/11/2020 - Peer reviewed on: 05/03/2021 - Reformulated on: 04/04/2021 - Recommended for publication on: 07/05/2021 - Published on: 30/06/2021

** Preprint of the article presented at XXIII SEMEAD, 2020.*