

Abordagem para Descoberta de Conhecimento em Instrumentos Avaliativos de Organizações de Ensino Superior

Leandro F. Lessa, Wladimir C. Brandão¹

¹Pontifícia Universidade Católica de Minas Gerais
Belo Horizonte – Minas Gerais – Brasil

leandrof.lessa@gmail.com, wladimir@pucminas.br

Abstract. *Performance evaluation is paramount to identify organizational weaknesses and strengths, providing issues to adjust strategies for increasing effectiveness. Several approaches to analyze data captured by evaluation instruments have been proposed in the literature and have been used to assess organizational performance. Particularly, in higher education organizations, such approaches are essential for self-criticism that leads to a continuous improvement of the organizational processes. This article proposes an approach to discover knowledge from evaluation instruments for higher education organizations. Content analysis and business intelligence techniques are used to automatically interpret answers to open and closed questions in evaluation instruments. The implementation of the proposed approach in a higher education organization demonstrated its effectiveness to support the knowledge discovery process for organizational performance.*

Resumo. *A condução de avaliações de desempenho é fundamental para a identificação de fraquezas e virtudes organizacionais, fornecendo importante subsídio para o ajuste de estratégias visando o aumento de efetividade. Diversas abordagens para análise de dados capturados por instrumentos avaliativos foram propostos na literatura e vêm sendo utilizados com o objetivo de avaliar o desempenho organizacional. Particularmente, em organizações de ensino superior tais abordagens são imprescindíveis para a formação de auto-crítica que as leve a uma melhoria contínua de processos. O presente artigo propõe uma abordagem para descoberta de conhecimento em instrumentos avaliativos de organizações de ensino superior. Técnicas de análise de conteúdo e de inteligência de negócio são utilizadas para interpretar automaticamente respostas a questões abertas e fechadas presentes nos instrumentos avaliativos. A aplicação da abordagem proposta no contexto de uma organização de ensino superior demonstrou a efetividade da abordagem para suportar o processo de descoberta de conhecimento relevante para o desempenho organizacional.*

1. Introdução

Existem diversos instrumentos de coleta de dados visando a avaliação do desempenho de organizações [Lakatos and Marconi 2010]. Esses instrumentos podem ser utilizados tanto para avaliações quantitativas, a fim de explicar relações causais, facilitando a generalização e prevendo acontecimentos através de formalismo e de medições estatísticas, quanto qualitativas, visando fornecer uma visão ampla de uma determinada situação, para compreender os processos sociais e suas interrelações [Cooper and Sghindler 2003].

A construção de questionários como instrumentos avaliativos consiste em produzir questionamentos claros e bem redigidos que traduzam os objetivos específicos da avaliação. Para que um questionário seja efetivo é fundamental a formulação de questões objetivas e que estejam intimamente relacionadas ao problema investigado [Gil 1999]. Do ponto de vista da forma, as questões podem ser abertas, onde as respostas devem ser dadas pelo respondente de maneira livre utilizando suas próprias palavras, ou fechadas, onde as respostas a serem dadas pelos respondentes são previamente estabelecidas pelos formuladores das questões.

A avaliação efetiva de uma organização pressupõe uma abordagem multidimensional com o objetivo de proporcionar uma compreensão abrangente de seus pontos fortes e fracos a partir dos resultados de avaliações individuais das pessoas que compõem a organização. Os dados relacionados à prática e ao desempenho individual são capturados a partir de diferentes pontos de vista, demandando habilidades analíticas especializadas e focadas em cada dimensão de análise definidas nos objetivos específicos da avaliação [Del Carmen Calatrava Moreno 2014].

O presente artigo propõe uma abordagem para descoberta de conhecimento em instrumentos avaliativos de organizações de ensino superior. Para interpretar respostas a questões abertas aplicamos técnicas de análise de conteúdo, um método para tratamento de informação contida em mensagens [Bardin 2009]. Para interpretar respostas a questões fechadas, aplicamos técnicas de inteligência de negócio, extraindo informação útil a partir dos dados capturados. A avaliação experimental da abordagem proposta foi efetuada a partir de sua aplicação no contexto de avaliação organizacional da Comissão Permanente de Avaliação (CPA) de um Instituto de Ensino Superior (IES)¹ e se mostrou efetiva para suportar a análise automática de questões abertas e fechadas de um questionário. Em particular, nossa abordagem possibilita que organizações possam extrair conhecimento de instrumentos de avaliação a fim de compreender e tratar suas fraquezas.

O conteúdo deste artigo está estruturado da seguinte maneira: a Seção 2 apresenta o referencial teórico, incluindo uma breve descrição das técnicas de inteligência de negócio e de análise de conteúdo. A Seção 3 descreve os trabalhos relacionados, em especial aqueles referenciados na literatura que propõem abordagens para descoberta de conhecimento em textos. A Seção 4, apresenta o fluxo de operação e os principais componentes da abordagem proposta. A Seção 5 descreve os resultados oriundos da aplicação da abordagem proposta no contexto da CPA de IES. Finalmente, a Seção 6 apresenta a conclusão, bem como direções para trabalhos futuros.

2. Referencial Teórico

2.1. Inteligência de Negócio

Um dos grandes desafios para as organizações é tomar decisões assertivas em tempo hábil baseadas na coleta de informação essencial e relevante disponível no mercado. Porém, com o grande volume de informação produzido pelas organizações, analisá-las torna-se cada vez mais complexo, uma vez que muitas das fontes de informação não a disponibiliza de maneira estruturada. Sendo assim, surge a necessidade de se criar plataformas de

¹A Pontifícia Universidade Católica de Minas Gerais, ou simplesmente PUC Minas, é o maior Instituto de Ensino Superior (IES) privado brasileiro.

inteligência de negócios, ou do inglês *business intelligence (BI)*, para viabilizar o processamento da informação sensível ao negócio em tempo real [Duan and Xu 2012]. Particularmente, por *BI* entende-se o processo de coleta, organização, análise, compartilhamento e monitoração de informação que oferece suporte à gestão de negócios [Batista 2004].

De maneira mais ampla, *BI* pode ser entendido como a utilização de várias fontes de informação para se traçar estratégias de competitividade para negócios da organização [Barbieri 2001]. Ela está relacionado ao apoio aos processos de tomadas de decisões baseados em dados trabalhados especificamente na busca de vantagens competitivas. Mais ainda, o *BI* é um sistema de informação que dá suporte à análise de dados do processo decisório, empregando tecnologias como *Data Warehouse*, *ETL*, *OLAP* e *Data Mining* nos níveis tático e estratégico [Audy 2005].

Técnicas de inteligência de negócio vêm sendo utilizadas em mecanismos de coleta de informação dentro de repositório de dados conhecidos como *Data Warehouses (DW)*. Tais repositórios são grandes bases de dados que contém dados históricos relativos às atividades de uma organização de forma consolidada. Em particular, um DW é um banco de dados histórico, separado em estruturas lógicas dimensionais, concebido para armazenar dados extraídos dos sistemas legados e do sistema de gestão integrada (ERP) da organização [Barbieri 2001]. Antes de serem armazenados no DW, os dados devem ser selecionados, organizados e integrados para que possam ser acessados de forma mais eficiente, auxiliando assim o processo de tomada de decisão [Colaço 2004].

O ETL, *Extract, Transformation and Load*, é a ferramenta responsável pela extração, transformação e carregamento dos dados em um DW. Na construção de um DW, o ETL consome mais de 70% do tempo de desenvolvimento, devido à diversidade da estrutura de dados nas bases de dados de origem [Colaço 2004]. O processamento analítico *online*, do inglês *Online Analytical Processing (OLAP)*, utiliza um conjunto de ferramentas e técnicas que permitem realizar a exploração dos dados de um DW, tais como recursos de modelagem, visualização e análise de grandes conjuntos de dados. O OLAP ajuda a analisar de forma eficiente a quantidade de dados armazenados pelas organizações transformando-os em informação relevante [Jacobson and Misner 2007].

2.2. Análise de Conteúdo

A análise de conteúdo pode ser definida como um conjunto de técnicas para análise de comunicação visando obter, através de procedimentos sistemáticos e objetivos de descrição do conteúdo das mensagens, indicadores que permitam a inferência de conhecimentos relativos às condições de produção e recepção dessas mensagens. Com essas técnicas, informação pode ser extraída a partir de dados, a fim de produzir novo conhecimento [Bardin 2009].

Para se aplicar de forma correta as técnicas de mineração de dados, faz-se necessário primeiramente a preparação dos dados para a análise de conteúdo, com o objetivo de identificar similaridades em função de morfologia ou de significado dos termos nos textos [Camilo and Silva 2009]. Mais especificamente, a preparação do texto envolve cinco etapas: coleta, pré-processamento, indexação, mineração e análise da informação. A etapa de coleta compõe a base de textos do trabalho. Para tal, é necessário determinar o universo de aplicação das técnicas de mineração de textos. Na etapa de pré-processamento é feito a transformação dos documentos no formato mais adequado a serem submetidos

aos algoritmos de extração de informação. Ela é responsável por obter uma representação estruturada dos documentos. Para isso, diversos algoritmos são aplicados a fim de melhorar a qualidade dos dados disponíveis e organizá-los [Oliveira 2009]. A etapa seguinte é a de indexação que envolve 3 processos: i) identificação dos termos, ii) remoção de *stopwords*, e iii) normalização morfológica ou (*stemming*). A identificação dos termos tem como objetivo principal a marcação dos termos contidos no texto, sejam eles simples ou compostos. Já a remoção *stopwords* implica em eliminar do texto original palavras consideradas não relevantes para a análise textual, tal como artigos, preposições, pronomes e advérbios. O *stemming* implica na redução das palavras a partir da eliminação de sufixos e prefixos das de variações morfológicas de uma palavra [Camilo and Silva 2009].

Após a indexação se inicia a etapa de processamento que consiste na aplicação de algoritmos de mineração responsáveis pelas inferências e cálculos buscando descobrir padrões úteis e desconhecidos presentes nos textos. Por fim, na etapa de análise é feita uma avaliação e interpretação dos resultados obtidos anteriormente [Oliveira 2009].

3. Trabalhos Relacionados

Esta seção apresenta os principais trabalhos relacionados à BI e análise de conteúdo. [Oliveira 2009] propõe uma metodologia para mineração de textos científicos em língua portuguesa. A autora apresenta as diversas etapas da mineração de textos com o intuito de agrupá-los segundo a similaridade de seus conteúdos de forma a facilitar a interpretação dos resultados e auxiliar na distribuição de documentos para determinados perfis de usuários. A autora concluiu que para realizar uma mineração de textos simples é necessário fazer um bom pré-processamento e para isso utilizar técnicas como remoção de *stopwords* e criação de um *Thesaurus*. Os melhores resultados foram alcançados a partir da combinação das análises estatísticas e semântica, combinação de mais de uma técnica de mineração de texto num mesmo contexto e a existência de um vocábulo controlado.

[Nooralahzadeh et al. 2013] realizaram a comparação dos sentimentos dos eleitores, antes e depois das eleições presidenciais, realizadas nos EUA e na França no ano de 2012. Para realizar essa comparação foram realizadas extrações das informações dos conteúdos a partir da rede social *Twitter*. O experimento teve o foco na polaridade nos sentimentos das pessoas (positivo e negativo) e nas palavras chaves que representava cada candidato. Primeiro foram extraídos *tweets* que continham informação dos candidatos. Em seguida foi realizada a limpeza de cada *tweet*, removendo *stopwords*, números, referências de HTML, símbolos, pontuações e *re-tweets*. Por fim, também foram extraídos os nomes dos candidatos de cada *tweet* e armazenados em um banco de dados, além da marcação dos *tweets* que não continham o nome dos candidatos. Esse passo foi realizado para atender a necessidade de futuras análises. Ao final do experimento, foi apresentada uma análise dos sentimentos, contendo a frequência das palavras mais citadas pelos eleitores e os resultados da popularidade dos presidentes avaliados.

[Ribeiro and Tostes 2015] os autores propuseram um sistema de inteligência de negócios para avaliação de organizações de ensino superior. O autor propôs uma metodologia de avaliação de organizações de ensino superior e uma plataforma Web de *Business Intelligence* para suportar a metodologia, automatizando parte do processo de avaliação institucional, e consequentemente auxiliando na geração de conhecimento estratégico. Diferente dos trabalhos supra apresentados, o presente trabalho utiliza técnicas de BI e de

análise de conteúdo para extrair conhecimento de instrumentos avaliativos de forma automatizada a partir da identificação e pesagem de descritores que irão permitir sumarizar o conhecimento presente nas respostas dos respondentes.

4. Abordagem para Descoberta de Conhecimento

A presença de questões abertas e fechadas em instrumentos avaliativos demanda a utilização de estratégias diferentes para tratamento de cada tipo de questão. Nesta seção apresentamos nossa abordagem para descoberta de conhecimento em instrumentos avaliativos. Particularmente, apresentamos as estratégias baseadas em análise de conteúdo e técnicas de inteligência de negócio para interpretação automática das respostas a questões apresentadas em instrumentos avaliativos. O fluxo do processo de descoberta de conhecimento tratado por nossa abordagem se encontra ilustrado na Figura 1.

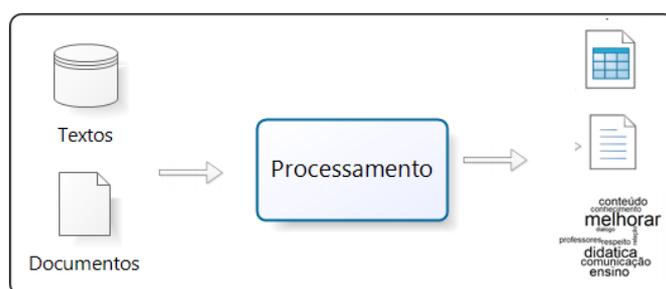


Figura 1. Fluxo de processo para questões abertas

4.1. Estabelecimento dos Parâmetros de Análise

A primeira etapa da nossa abordagem consiste em receber os documentos e o texto em linguagem natural referente as respostas dos avaliadores. Os documentos por sua vez, são responsáveis por guiar a avaliação e possuem dimensões e indicadores já estabelecidos antes da aplicação dos questionários. Na segunda etapa é realizado o processamento dos dados coletados. Os textos são extraídos de um banco de dados utilizado na coleta da avaliação. Após isso, os dados são tratados de acordo com as especificações dos documentos. Por fim na última etapa, temos o produto de cada processamento dos dados. Nesta abordagem obtivemos a categorização das perguntas e respostas, classificação de sentimentos das respostas, definição do vocábulo controlado e uma nuvem de TAGs com a frequência das palavras mais ditas pelos entrevistados.

4.2. Processamento de Questões Abertas

Questões abertas são compostas por um texto em linguagem natural, geralmente uma pergunta, que demanda uma resposta também em linguagem natural por parte do respondente. Esse tipo de questionamento visa o aprofundamento do conhecimento sobre o tema investigado a partir da captura da percepção do respondente, com foco na extração de suas opiniões [Richardson and Jarry 1985]. Tal como ilustrado na Figura 2, nossa abordagem subdivide o processamento de questões abertas em cinco etapas: i) definir categoria da pergunta; ii) definir categoria da resposta; iii) classificação das respostas através de qualificadores; iv) criação do vocabulário controlado; v) criação de nuvem de TAGs.

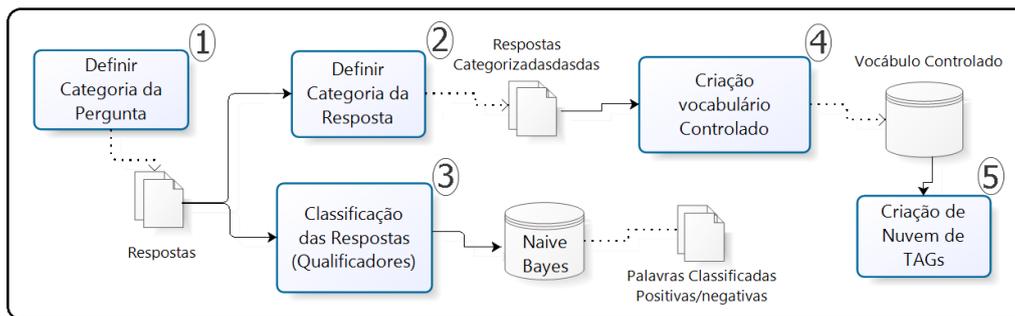


Figura 2. Etapas da abordagem para questões abertas

Definir categoria da pergunta: O primeiro passo é categorizar a pergunta de acordo com o conteúdo. As perguntas podem ser categorizadas de acordo com seu conteúdo da seguinte forma: (a) perguntas sobre fatos; (b) perguntas sobre crenças; (c) perguntas sobre sentimentos; (d) perguntas sobre padrões de ação (e) perguntas dirigidas a comportamento presente ou passado; (f) perguntas referentes a razões conscientes de crenças, sentimentos, orientações ou comportamentos [Gil 1999].

Definir categoria da resposta: O segundo passo é identificar as categorias dessa resposta. Uma característica de uma pergunta aberta de um questionário é a possibilidade de o entrevistado responder com mais liberdade, não estando restrito a marcar alternativas previamente elaboradas. Além de fornecer ao entrevistador informações a respeito do entrevistado sobre um determinado assunto, possibilita também a possibilidade de expressar sentimentos [Richardson and Jarry 1985].

Classificação das respostas através de qualificadores: O terceiro passo consiste em avaliar um número de respostas da amostra, definindo de forma manual se a resposta é considerada positiva ou negativa. Após isso é necessário criar qualificadores que serão utilizados para realizar análise de sentimentos das respostas fornecidas pelos entrevistados. Os qualificadores são um conjunto de palavras com polaridade positivas e negativas que são utilizados de base para análise de cada categoria de avaliação. Baseado nos dados colhidos na amostra e nos qualificadores, o algoritmo *Naive Bayes* consegue inferir de forma automática as demais tendências de respostas fornecidas pelo demais entrevistados.

Criação do vocabulário controlado: O quarto passo é a criação do vocabulário Controlado. Que é uma linguagem artificial constituída de termos de uma estrutura relacional. Seu objetivo é padronizar e facilitar a entrada e saída de dados de um sistema de informação. Uma das funções é representar a informação e o conhecimento por meio de um conjunto controlado e finito de termos.

Criação de nuvem de TAGs: Existem diversas formas de classificar e organizar informações de um texto. Uma delas é a forma representativa em nuvem de TAGs. Que consiste mostrar visualmente através do tamanho das palavras os termos mais utilizado em um texto. O quinto e último passo da abordagem é criar uma nuvem de TAGs. Para isso é necessário identificar todos os termos que possuem maior frequência nas respostas por categoria e representá-los.

4.3. Processamento de Questões Fechadas

Questões fechadas são compostas por um texto em linguagem natural, geralmente uma pergunta, que demanda uma resposta com opções limitadas a alternativas preestabelecidas por parte do respondente. Esse tipo de questionamento visa o aprofundamento do conhecimento sobre o tema investigado a partir da captura da percepção do respondente, com foco na extração de suas opiniões [Richardson and Jarry 1985]. Tal como ilustrado na Figura 3, nossa abordagem subdivide o processamento de questões fechadas em quatro etapas: i) análise dos dados; ii) criação do DW; iii) criação de processo ETL; iv) criação de relatórios.

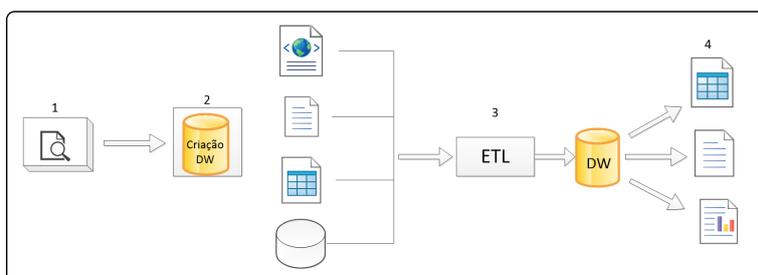


Figura 3. Etapas da abordagem para questões fechadas

Etapa 1 - Análise dos dados: Antes da criação um DW, é muito importante conhecer o banco de dados que será utilizado e as tabelas que relacionam entre si. Para isso é necessário fazer um mapeamento completo e detalhado de todo do modelo, afim de identificar a fonte dos dados que o alimenta e conhecer cada detalhe. Nessa etapa é feito o levantamento de todos os dados que serão utilizados e exibidos no modelo de DW.

Etapa 2 - Criação do DW: Definido os mapeamentos, os bancos e tabelas a serem utilizados, o próximo passo é criar o esquema de DW. O esquema adotado foi o *Star Schema* proposto por Dr. Ralph Kimball e que tem como característica básica a presença de dados altamente redundantes para se obter um melhor desempenho [Colaço 2004].

Etapa 3 - Criação de processo ETL: Esse processo é o mais crítico e demorado na construção de um DW, pois consiste na extração dos dados de bases heterogêneas, na transformação e limpeza e carga desses dados.

Etapa 4 - Criação de relatórios O último passo é a etapa de criação dos relatórios. Essa etapa é o produto final da análise. Aqui podemos criar diversos tipos de relatórios que irá exibir o conhecimento extraído do DW.

5. Avaliação da Abordagem Proposta

Nessa seção vamos demonstrar os resultados obtidos na proposta de avaliação para questões abertas e fechadas aplicadas no questionário. Para questões abertas realizamos um estudo de caso com uma questão e por motivo de sigilo institucional, não será revelada. Já para as questões fechadas, optamos por criar uma plataforma de BI para análise das questões, bem como criação de relatórios que auxiliam no processo de decisão estratégias.

5.1. Análise das Questões Abertas

Definição da categoria da pergunta: A questão selecionada e analisada tinha objetivo obter as reações emocionais dos entrevistados referente a um curso ofertado da instituição.

Neste caso, a questão foi categorizada como pergunta sobre sentimentos.

Definição da categoria da resposta: Após definida a categoria da pergunta, o próximo passo é definir a categoria da resposta. Na nossa abordagem optamos desenvolver um aplicativo na linguagem C# que realiza a categorização automática das respostas dos entrevistados. Para isso, foi criado um dicionário de categorias que auxiliou na categorização. Um especialista no assunto criou as seguintes categorias de respostas: i) infraestrutura, ii) projeto pedagógico, iii) relações ente os sujeitos, iv) condições dos alunos, v) condições dos professores, vi) condições da coordenação.

Classificação das respostas através de qualificadores: O próximo passo da abordagem é inferir os sentimentos dos entrevistados a partir do algoritmo *Naive Bayes*. Para isso, foi necessário fazer o uso de uma amostra de cinquenta respostas classificadas manualmente e criar um arcabouço de palavras com polaridades positivas e negativas que foram utilizadas no treinamento do algoritmo. A Tabela 1 demonstra uma prévia das palavras que tem polaridade negativa e positiva com seus respectivos pesos inferidas pelo algoritmo.

Tabela 1. Polarização das Palavras

Palavras com polaridade negativa		Palavras com polaridade positiva	
chato	3	alegre	5
horrrível	5	empolgante	5
desagradável	1	bom	4
desinteressante	2	brilhante	5
monótono	5	exemplar	4

A Tabela 2 mostra a classificação automática realizada e o percentual de respostas com polaridade positivas e negativas. Pode ser observado que existem respostas que estão categorizadas em mais de uma categoria. Isso já era esperado, tratando-se de questões abertas de um questionário, as respostas dos entrevistados podem se adequar a uma ou de mais categorias. Grande parte das respostas dos entrevistados foram classificadas como respostas com polaridade negativas. Isso aconteceu devido a questão não ter sido preparada anteriormente para realizar esse tipo avaliação. O objetivo da questão era levantar informações de forma geral problemas de um determinado curso da instituição.

Tabela 2. Respostas Positivas e Negativas por Categoria

Categoria da avaliação	Nº de Respostas da Categoria	Respostas Positivas	Respostas Negativas
Infraestrutura	10	2,0%	98,0%
Projeto Pedagógico	30	1,8%	98,2%
Relação entre os sujeitos	60	2,2%	97,8
Condições dos professores	150	1,2%	98,8%
Condições dos professores e coordenação	189	1,3%	98,7%
Condições da coordenação	120	0,2%	99,8%
Infraestrutura e projeto pedagógico	30	3,0%	97,0%

Criação do vocabulário controlado: O vocabulário controlado é um recurso utilizado para criar palavras chaves que caracterizam uma determinada categoria de resposta. A Tabela 3 mostra uma prévia do vocabulário controlado criado por um especialista para a categoria de infraestrutura.

Tabela 3. Vocabulário Controlado de Infraestrutura

Palavras		
apagador	laboratório	projedor
ar condicionado	mesa	quadro
cadeira	pincel	sala de aula

Criação de nuvem de TAGs: As nuvens de TAGs é o resultado da classificação automática criada pelo aplicativo desenvolvido para avaliar as respostas do questionário de questões abertas. A Figura 4 apresenta as palavras com maior frequência respondidas pelos entrevistados referente a condições dos professores do curso avaliado.



Figura 4. Frequência das Palavras Referente a Condições dos Professores

Para o cenário de avaliação da questão analisada, observamos um sentimento negativo em relação condições dos professores, isso indica que deve-se melhorar os principais tópicos destacados na nuvem de TAGs. Esses tópicos podem mudar de acordo com as respostas dos alunos.

5.2. Análise das Questões Fechadas

Análise dos dados: O primeiro passo da abordagem proposta é o estudo dos dados. Por se tratar de uma etapa fundamental para criação de uma plataforma BI, foi necessário realizar um estudo detalhado do dados junto a equipe de infraestrutura e banco de dados da CPA. Esse estudo foi feito através de reuniões que proporcionou o entendimento das demandas da instituição, bem como, realizar o mapeamento, levantamento e identificação dos dados que seriam analisados.

Criação do DW para a CPA de IES: Após o estudo dos dados, foi desenvolvido o modelo do DW. O esquema adotado nesse trabalho é composto por uma tabela dominante, chamada de tabela fato, no centro, rodeada por diversas tabelas auxiliares, chamadas de tabelas dimensões. A tabela fato se comunica com as demais tabelas dimensões através de múltiplas junções e as tabelas dimensões se conecta apenas por uma única junção a tabela fato. A Figura 5 ilustra o modelo desenvolvido pela abordagem.

Nesse esquema podemos realizar a pesquisa inicialmente pelas tabelas dimensões e depois na fato, isso garante um acesso mais eficiente e um desempenho alto, visto que não é necessário percorrer todas as tabelas para ter acesso aos dados.

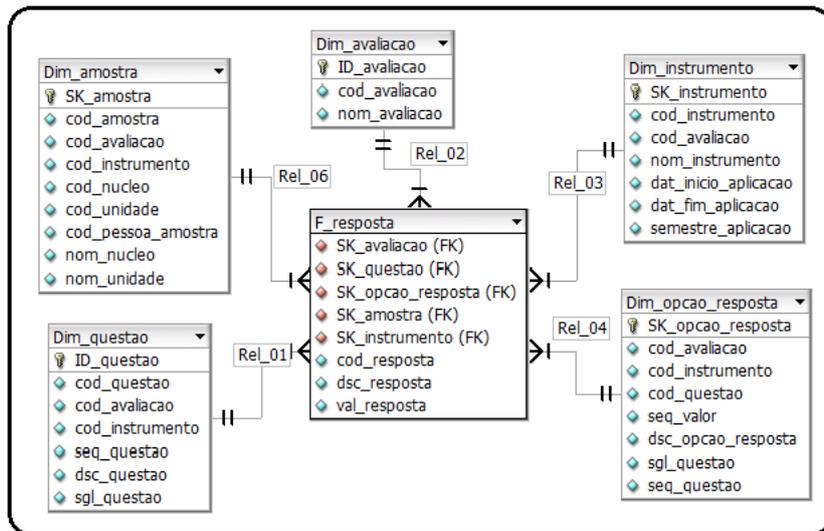


Figura 5. DW – ESQUEMA ESTRELA

Criação de processos ETL: Após a criação do esquema do DW no banco de dados, o próximo passo foi criar processos ETLs na ferramenta Microsoft SQL Server Integration Services. Foram criados dois processos ETL distintos. O primeiro para dar carga nas tabelas dimensões e o segundo para a tabela fato. O processo ETL de carga nas tabelas dimensões pode ser visto através da Figura 6.

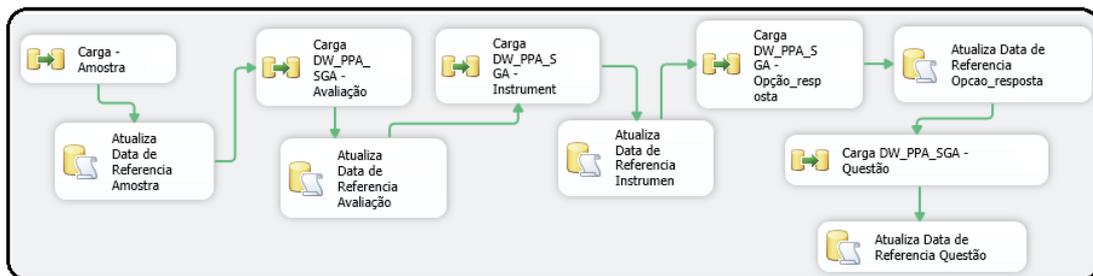


Figura 6. Processo de Carga ETL

Resultados da criação do DW: Um experimento foi realizado com um banco de dados da CPA, com o objetivo de realizar uma consulta através dos questionários de questões fechadas. Conforme mostra a Figura 7 podemos observar como os dados são agrupados através de níveis e hierarquias. Para uma análise mais sucinta esse agrupamento é mais adequado e facilita análise dos dados. Nesse experimento é demonstrado que existem 3 tipos de avaliação, no qual 67,59% das pessoas entrevistadas responderam ao questionário de avaliação institucional 2013.

As consultas realizadas em um DW podem ser detalhadas a um nível mais baixo de informação. Conforme descemos o nível de detalhamento podemos descobrir dados que antes eram ocultos e superficiais passam a ser relevantes e significativos para a análise de conhecimento. A Figura 8 detalha os níveis e hierarquia expandidos entre as informações referentes ao questionário de engenharia, onde podemos explorar ao nível de questões e dados referentes as suas respostas.

Nom Avaliacao	Frequencia
Avaliação Docente	0,19%
Avaliação Institucional_2013	67,59%
Questionário Engenharia	32,23%
Grand Total	100,00%

Figura 7. Consulta da Base de Dados da CPA (Alto Nível)

Nom Avaliacao	Nom Instrumento	Dsc Questao	Dsc Opcao Resposta	Frequencia
Avaliação Docente				0,19%
Avaliação Institucional_2013				67,59%
Questionário Engenharia	Questionári Engenharia	Aluno matriculado na disciplina Trabalho de Conclusão de Curso		0,56%
		Atitudes Outra. Qual?		0,56%
		Atitudes Procuro a monitoria.	Não adoto essa atitude	55,33%
			Não respondeu	2,43%
			Não sou atendido	9,16%
			Sou atendido com frequência	12,52%
			Sou atendido de vez em quando	20,56%
			Total	100,00%

Figura 8. Consulta da Base de Dados da CPA (Baixo Nível)

Através de um esquema DW podemos realizar diversos relatórios de forma rápida e dinâmica de acordo com a demanda da instituição. A Tabela 4 mostra um relatório produzido pela CPA referente a pergunta aplicada em um questionário. A pergunta realizada foi: “Você desenvolve alguma atividade remunerada (inclusive estágio remunerado)?”

Tabela 4. Relatório Aplicado CPA

Opções de Respostas	Nº Respostas	%
Sim, trabalho até 10 horas semanais	151	6.8
Sim, trabalho de 10 a 20 horas semanais	184	8.2
Sim, trabalho de 20 a 30 horas semanais	281	12.6
Sim, trabalho entre 30 e 40 horas semanais	229	10.3
Sim, trabalho mais de 40 horas semanais	165	7.4
Não estou trabalhando no momento	1221	54.7
Total	2231	100

Através desse relatório, podemos observar que a maioria dos entrevistados responderam que não estão trabalhando no momento o que representa 54,7% das respostas. O uso de um DW para criação de relatórios se torna uma ferramenta poderosa nas mãos das instituições para geração de conhecimento.

6. Conclusão

O presente artigo propôs uma abordagem para descoberta de conhecimento em instrumentos avaliativos de organizações de ensino superior. A abordagem proposta foi aplicada no contexto de avaliação organizacional da Comissão Permanente de Avaliação (CPA) de IES e se mostrou efetiva para suportar a análise automática de questões abertas e fechadas. Particularmente, técnicas de análise de conteúdo foram utilizadas para interpretar respostas a questões abertas, enquanto técnicas de inteligência de negócio foram utilizadas para interpretar respostas a questões fechadas, fornecendo um arcabouço ferramental

que permitiu a descoberta de conhecimento relevante sobre os pontos fortes e fracos do desempenho organizacional.

Como trabalhos futuros pretendemos incorporar em nossa abordagem as funcionalidades disponibilizadas pela plataforma de BI desenvolvida em SharePoint tal como descrito em [Ribeiro and Tostes 2015]. Além disso, pretendemos estender a avaliação da abordagem proposta, aplicando-a em outras organizações de ensino superior e em outros contextos organizacionais.

Referências

- Audy, J. L. N.; Andrade, G. K. C. (2005). *Fundamentos de sistemas de informação*. Porto Alegre: Bookman.
- Barbieri, C. (2001). *BI-business intelligence: modelagem e tecnologia*. [S.l.]: ISBN 9788573231489.
- Bardin, L. (2009). *Análise de Conteúdo*. Portugal; Edições 70, LDA.
- Batista, E. d. O. (2004). *Sistema de Informação : o uso consciente da tecnologia para o gerenciamento*. São Paulo.
- Camilo, O. C. and Silva, J. C. (2009). Mineração de dados: Conceitos, tarefas, métodos e ferramentas. Master's thesis, UFG.
- Colaço, M. (2004). *Projetando sistemas de apoio à decisão baseados em data warehouse*. ISBN: 8573232080.
- Cooper, D. R. and Sghindler, P. S. (2003). *Business Research Methods*.
- Del Carmen Calatrava Moreno, M. (2014). A 360-degree evaluation framework for doctoral programs. *IEEE Global Engineering Education Conference (EDUCON)*, pages 850 – 853.
- Duan, L. and Xu, L. D. (2012). Business intelligence for enterprise systems: A survey. *Industrial Informatics, IEEE Transactions on*, 8:679 –687.
- Gil, A. C. (1999). *Métodos e técnicas de social*. Capítulo II. O questionário. Conceituação. Vantagens e limitações do questionário. A construção do questionário. p. 124-132.
- Jacobson, R. and Misner, S. (2007). *Microsoft: SQL Server 2005 Analysis Services*. ISBN 9788560031719.
- Lakatos, E. and Marconi, M. (2010). *Fundamentos de metodologia científica*. ISBN 9788522457588.
- Nooralahzadeh, F., Viswanathan, A., and Costin, C. (2013). 2012 presidential elections on twitter - an analysis of how the us and french election were reflected in tweets. *International Conference on Control Systems and Computer Science*, page 7.
- Oliveira, I. M. (2009). Estudo de uma metodologia de mineração de textos científicos em língua portuguesa. Master's thesis, COPPE/UFRJ.
- Ribeiro, C. E. and Tostes, A. I. J. T. R. (2015). Metodologia de geração de plataforma de inteligência de negócios para comissões de avaliação de universidade.
- Richardson and Jarry, R. (1985). *Pesquisa social; métodos e técnicas*. Capítulo 9. Questionário. Funções e características. p. 142-150.