



INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA DO PIAUÍ
CAMPUS TERESINA CENTRAL
TECNOLOGIA EM ANÁLISE E DESENVOLVIMENTO DE SISTEMAS

NAYARA DA SILVA FERREIRA

LISE - ANÁLISE DE CONCORRENTES AUTOMATIZADA

TERESINA, PIAUÍ

2018

NAYARA DA SILVA FERREIRA

LISE - ANÁLISE DE CONCORRENTES AUTOMATIZADA

Projeto apresentado à Banca Examinadora como requisito para aprovação na disciplina de Trabalho de Conclusão de Curso II do Curso Superior de Tecnologia em Análise e Desenvolvimento de Sistemas do Instituto Federal de Educação, Ciência e Tecnologia do Piauí.

Orientador: Prof. M^º. Valéria Oliveira Costa

TERESINA, PIAUÍ

2018

FICHA CATALOGRÁFICA

Sistema de Bibliotecas

Gerada automaticamente com dados fornecidos pelo(a) autor(a)

F345I Ferreira, Nayara da Silva
Lise - análise de concorrentes automatizada / Nayara da Silva Ferreira - 2018.
60 p. : il. color.

Trabalho de conclusão de curso (Tecnologia) - Instituto Federal de Educação,
Ciência e Tecnologia do Piauí, Campus Teresina Central, Tecnologia de Alimentos, 2018.

Orientadora : Profa Ma. Valéria Oliveira Costa.

1. Análise de concorrentes. 2. Mineração de dados. 3. Plano de negócio. 4. Google
Maps. 5. Empreendedor. I.Título.

CDD 664

FOLHA DE APROVAÇÃO

TÍTULO: LISE- ANÁLISE DE CONCORRENTES AUTOMATIZADA

ALUNA: NAYARA DA SILVA FERREIRA

DATA: 14 de março de 2018

Trabalho de conclusão da Disciplina TCC II, defendido junto ao Departamento de Informação, Ambiente, Saúde e Produção Alimentícia do Instituto Federal de Educação, Ciência e Tecnologia do Piauí.

BANCA EXAMINADORA

Assinatura no original

Prof^a. Msc. Valéria Oliveira Costa
Instituto Federal do Piauí (IFPI)
Orientador

Assinatura no original

Prof. Msc. Ely da Silva Miranda
Instituto Federal do Piauí (IFPI)
Avaliador

Assinatura no original

Prof^a. Msc. Adalton de Sena Almeida
Instituto Federal do Piauí (IFPI)
Avaliadora

A todos que estiveram próximos de mim, pelo incentivo e pelo apoio constante.

AGRADECIMENTOS

Agradeço primeiramente a meus pais, minha irmã e Diego Bruno pelo apoio, incentivo e confiança nas minhas decisões.

Agradeço aos colegas de estudos do curso e das bibliotecas que me ajudaram bastante. Em especial aqueles que estiveram comigo desde o início dessa jornada.

Agradeço profundamente aos mestres e professores que compartilharam de seus conhecimentos por todos esses anos.

Por fim, agradeço a minha orientadora, Valéria Oliveira, por sua amizade, motivações, conselhos e ótimas orientações que fizeram este trabalho caminhar.

“Determinação, coragem e auto-confiança são fatores decisivos para o sucesso. Se estamos possuídos por uma inabalável determinação, conseguiremos superá-los. Independentemente das circunstâncias, devemos ser sempre humildes, recatados e despidos de orgulho.”
(Dalai Lama, tradução livre)

RESUMO

Hoje em dia, qualquer pessoa com uma ideia inicial pode se tornar um empreendedor. Especialistas recomendam aos empreendedores um planejamento da ideia antes de aplicá-la no mercado. Sendo uma das etapas do plano de negócios, a análise de concorrentes ajuda o empreendedor a criar estratégias para ter vantagens no mercado. Porém, o processo manual de análise de concorrentes é demorado e cansativo, podendo desanimar o empreendedor a realizar esta etapa.

Para minimizar esse problema existem dois tipos de ferramentas computacionais que o empreendedor pode utilizar: ferramentas de pesquisa com análise automática e ferramentas de pesquisa com visão ampla no mercado. O primeiro tipo exige participações de várias pessoas envolvidas com os concorrentes para responder questionários. Por exemplo, as ferramentas *Survio*, *Google Forms* e *SurveyMonkey*. O segundo tipo coleta dados da Internet tendo uma visão ampla do mercado para realizar análise automática, como *Alexa*, *Open Site Explorer* e *SemRush*. Neste último caso, há um problema para empreendedores de empresa locais, pois o ideal é analisar suas proximidades.

Este trabalho apresenta como solução o Lise, um sistema desenvolvido para automatizar o processo de análise de concorrentes, em forma de uma plataforma *web*, utilizando o *Google Maps* para coletar dados dos concorrentes da localidade que o empreendedor deseja desenvolver sua empresa. Como contribuição esperada, este trabalho pretende que o Lise ajude os empreendedores analisarem seus concorrentes de forma rápida e fácil.

Palavras-chaves: análise de concorrentes. empreendedores. ferramenta computacional. plano de negócios. google maps.

ABSTRACT

Nowadays, anyone with an initial idea can become an entrepreneur. Experts advise entrepreneurs to plan the idea before applying it to the market. Being one of the steps in the business plan, competitor analysis helps the entrepreneur devise strategies to take advantage of the market. However, the manual process of analyzing competitors is time-consuming and tiresome, and can discourage the entrepreneur to perform this step.

To minimize this problem there are two types of computational tools that the entrepreneur can use: research tools with automatic analysis and research tools with wide vision in the market. In the first type requires the participation of several people involved with competitors to respond to questionnaires. For example, the Survio, Google Forms and SurveyMonkey tools. The second type collects data from the Internet taking a broad view of the market to perform automatic analysis such as Alexa, Open Site Explorer and SemRush. In the latter case, there is a problem for local entrepreneurs, because the ideal is to analyze their vicinity.

This paper presents as a solution the Lise, a system developed to automate the process of competitors analysis, in the form of a web platform, using Google Maps to collect data of the competitors of the locality that the entrepreneur wants to develop his company. As an expected contribution, this paper intends that Lise helps entrepreneurs analyze their competitors quickly and easily.

Key-words: competitors analysis. entrepreneurs. computational tool. business plan. google maps.

LISTA DE ILUSTRAÇÕES

Figura 1 – Hierarquia dos objetos	26
Figura 2 – Diagrama de casos de uso	28
Figura 3 – Diagrama UML de classes	29
Figura 4 – Arquitetura do Lise	31
Figura 5 – Arquitetura do módulo de gerenciamento de dados	32
Figura 6 – Resultado da polarização da verificação	37
Figura 7 – Resultado da sumarização da verificação	37
Figura 8 – Página de cadastro do plano de negócio do Lise	38
Figura 9 – Lista de planos de negócios do empreendedor do Lise	39
Figura 10 – Modelos dos tipos de informações produzidas	41
Figura 11 – Página de gerenciamento de usuários	42
Figura 12 – Página de gerenciamento de negócios planejados	42
Figura 13 – Página de gerenciamento de categorias	43
Figura 14 – Página de gerenciamento de solicitações de categorias	43
Figura 15 – Página de gerenciamento de atualização de dados	44
Figura 16 – Página de gerenciamento de tarefas	44

LISTA DE TABELAS

Tabela 1 – Critérios para teste de hipótese	25
Tabela 2 – Requisitos funcionais e não funcionais do Lise	27
Tabela 3 – Exemplo de resultados das fases de gerenciamento de dados	33
Tabela 4 – Categorias e localidades utilizadas na verificação	35
Tabela 5 – Tópicos empresariais para basear na sumarização	36
Tabela 6 – Pontos analisados dos tradutores	51

LISTA DE ABREVIATURAS E SIGLAS

API	Application Programming Interface
CoGrOO	Corretor Gramatical para OpenOffice
GNMT	Google Neural Machine Translation
HTML	HyperText Markup Language
NMT	Neural Machine Translation
REST	Representational State Transfer
SMT	Statistical Machine Translation
UML	Unified Modeling Language

SUMÁRIO

1	INTRODUÇÃO	14
1.1	Contextualização	14
1.2	Justificativa	14
1.3	Objetivos	15
1.3.1	Objetivo Geral	15
1.3.2	Objetivos Específicos	15
1.4	Organização do Texto	16
2	REFERENCIAL TEÓRICO	17
2.1	Análise de Concorrentes	17
2.2	Recuperação de Informação	17
2.3	Mineração de Dados	18
2.4	Análise de Sentimento	18
2.5	Ferramentas de Apoio aos Empreendedores	19
3	METODOLOGIA	21
3.1	Recursos Tecnológicos	21
3.2	Base de Dados	22
3.3	Verificação do Sistema	23
4	LISE	26
4.1	Nomenclaturas dos Objetos	26
4.2	Requisitos	27
4.3	Diagramas	28
4.3.1	Casos de Uso	28
4.3.2	Diagrama de Classes	28
4.4	Arquitetura	31
4.4.1	Módulo de Gerenciamento de Dados	32
4.4.2	Módulo de Aplicação	33
5	RESULTADOS	35
5.1	Verificação	35
5.2	Interface da Aplicação	38
6	CONCLUSÃO	46
6.1	Trabalhos Futuros	46

REFERÊNCIAS	48
APÊNDICE A – SELEÇÃO DE TRADUTOR PARA O LISE	51
APÊNDICE B – ESPECIFICAÇÃO DE CASOS DE USO	52
APÊNDICE C – LISTA DE CATEGORIAS INICIAL DO LISE	56
ANEXO A – TABELA DE DISTRIBUIÇÃO NORMAL PADRÃO	59

1 INTRODUÇÃO

1.1 CONTEXTUALIZAÇÃO

Nos últimos três anos, o número de pessoas que se tornam empreendedores cresceu constantemente, por oportunidade ou necessidade (MELO, 2017). O empreendedor é a pessoa que inicia e dinamiza um negócio para realizar uma ideia ou projeto pessoal assumindo riscos e responsabilidades (CHIAVENATO, 2004). Para abrir uma empresa, é fundamental que os empreendedores produzam um plano de negócios. Segundo Dornelas (2013), o plano de negócios é um documento utilizado para planejar um empreendimento, em estágio inicial ou não, com o propósito de definir e delinear sua estratégia de atuação para o futuro.

Uma das principais etapas do plano de negócios é a análise de concorrentes (GIBICOSKI, 2017). Esta análise capacita o empreendedor a compreender melhor a situação do mercado, além de apontar os pontos fracos dos concorrentes para que possam ser aperfeiçoados (EGESTOR, 2017). Rosa (2004) explicou um método de fazer a análise de concorrentes manual. Neste método o empreendedor deve observar a atuação da concorrência, identificar quem são os seus principais concorrentes, depois visitá-los e examiná-los procurando as boas práticas e deficiências. Com uma tabela, o empreendedor deverá enumerar os pontos fortes e fracos de cada concorrente para poder comparar estes pontos e tirar conclusões da pesquisa. Com base neste método, o Lise realiza a análise de concorrentes, porém de forma automatizada.

Com o passar do tempo as redes sociais se tornaram uma ferramenta de divulgação de opiniões, inclusive sobre as empresas. Também, crescentemente, pequenas e grandes empresas estão se conectando às redes sociais, seja com seu próprio perfil em algumas, como as *fanpages* do *Facebook*¹, ou sendo objetos dos serviços oferecidos em outras. Um exemplo deste segundo caso é o *Google Maps*². Este serviço do *Google* mostra várias informações das empresas cadastradas nela. Através dele, os seus usuários podem localizar um determinado estabelecimento, encontrar contatos, visualizar avaliações de outros clientes, se informar sobre o horário de serviço, entre diversas outras utilidades. Esta aplicação do *Google* proporciona vantagens tanto para os estabelecimentos, que ficam mais conhecidos e fáceis de localizar, quanto para os consumidores, que podem visualizar detalhes do local sem precisarem sair de casa.

1.2 JUSTIFICATIVA

Em diversas análises, basicamente, os dados são os principais elementos e quando são processados geram informações. Para a análise de concorrentes estes dados não são encontrados facilmente. Geralmente, este processo é realizado manualmente utilizando pesquisas extenuantes.

¹ <https://www.facebook.com/>

² <https://www.google.com.br/maps/>

O empreendedor gastaria muito tempo somente para coletar poucos dados. Pelo fato de serem cansativas, estas pesquisas podem desanimar o empreendedor e fazê-lo não concluir a etapa de análise de concorrentes ou produzir resultados de qualquer maneira para prosseguir. Extrair este conteúdo da internet pode ser uma maneira de agilizar o processo de coleta.

Outro problema da análise de concorrentes manual é organizar os dados coletados para, em seguida, ter uma boa análise. Por exemplo, quanto trabalho um pequeno empreendedor teria para classificar as respostas de vários questionários? E para selecionar os melhores ou os piores pontos dos seus concorrentes com base nas opiniões sobre outras empresas do mesmo ramo empresarial? Nestas situações, o tempo gasto somente neste processo manual pode ser inviável. Com estas circunstâncias, o processo automatizado também agilizaria a organização.

Planejar um negócio proporciona uma visão mais clara sobre as estratégias do empreendedor e como ele espera alcançá-las. O planejamento não dá a certeza do sucesso, mas se um negócio for mal planejado, há grandes chances de acarretar prejuízos e até fechamento da empresa. Por isso, o Sebrae considerou o planejamento de negócios como um dos fatores contribuintes para a sobrevivência das empresas (MICRO, 2013). O uso de uma ferramenta automatizada de análise de concorrentes poderá motivar essas pessoas e ajudá-las a seguir o processo com uma visão mais clara do mercado.

A ideia principal do Lise é utilizar as avaliações dadas pelos clientes dos estabelecimentos através do *Google Maps* para encontrar pontos positivos e negativos da concorrência de uma determinada categoria de negócio. Isso porque é na avaliação que o cliente coloca suas opiniões, reclamações ou elogios de acordo com a experiência que teve. Além disso, o *Google* proporciona um sistema de recompensas para motivar, cada vez mais, os usuários avaliarem os estabelecimentos e assim gerarem *feedbacks* para as empresas cadastradas no *Google Maps* (KURTZ, 2017). Além desta principal funcionalidade, serão utilizados dados de cadastros dos estabelecimentos do *Google Maps* para gerarem informações úteis sobre os concorrentes como horários de funcionamento, regiões dos concorrentes na cidade, entre outras.

1.3 OBJETIVOS

1.3.1 Objetivo Geral

Desenvolver uma ferramenta de apoio aos empreendedores de empresas físicas para análise de concorrentes com base nos dados públicos do Google Maps.

1.3.2 Objetivos Específicos

Como contribuintes para alcançar o objetivo principal, foram definidos os seguintes objetivos específicos:

- Coletar dados públicos para gerar as análises de concorrentes;

- Disponibilizar as informações das análises de concorrentes geradas no Lise;
- Facilitar e agilizar a análise de concorrentes para os empreendedores.

1.4 ORGANIZAÇÃO DO TEXTO

Este documento está dividido em seis capítulos. Neste primeiro capítulo foi citada a contextualização do tema escolhido e a justificativa do projeto, assim como os objetivos gerais e específicos. Os próximos capítulos são organizados na seguinte forma:

- **Referencial teórico:** Este capítulo apresenta um maior detalhamento sobre análise de concorrentes e aborda outros conceitos utilizados para o desenvolvimento deste projeto.
- **Metodologia:** O capítulo de metodologia propõe de forma ampla o método de desenvolvimento do projeto, destacando a base de dados. Além disso, este capítulo descreve a verificação dos resultados gerados pelo Lise.
- **Lise:** No capítulo da ferramenta estão descritos os seus requisitos, os principais diagramas, a arquitetura do sistema e as etapas dos módulos arquitetados.
- **Resultados:** Este capítulo mostra os resultados da verificação e a interface das informações geradas pelo Lise.
- **Conclusão:** O último capítulo conclui este trabalho e descreve alguns trabalhos futuros.

RESUMO

Neste capítulo, inicialmente, foi feita uma contextualização sobre plano de negócios e análise de concorrentes, falando um pouco sobre o processo manual desta análise e mencionando o funcionamento do *Google Maps*, a fonte de dados do Lise. Em seguida, foi mostrada a justificativa e os objetivos que motivaram este trabalho. Por fim, a organização deste texto foi descrita.

2 REFERENCIAL TEÓRICO

Alguns conceitos são importantes para uma melhor compreensão deste trabalho. Dentre eles, os principais são: análise de concorrentes, recuperação de informação, mineração de dados e análise de sentimento. Além destes conceitos, é importante conhecer algumas ferramentas que o empreendedor poderia utilizar como apoio para análise de concorrentes, afim de compreender ainda mais a utilidade e o diferencial desse projeto.

2.1 ANÁLISE DE CONCORRENTES

Como foi dito no capítulo de introdução, a análise de concorrentes é uma etapa do plano de negócios. Por isso é importante entender alguns conceitos de plano de negócios antes de tratar sobre a análise de concorrentes. Dornelas (2011) definiu o plano de negócios como um documento para planejar estratégias do futuro de um empreendimento em estágio inicial ou não. Segundo o autor, o plano de negócios serve para testar a viabilidade de um conceito de negócio, orientar o desenvolvimento da estratégia e de operações, atrair recursos financeiros, transmitir credibilidade e desenvolver a equipe de gestão. Para Casarotto (2009), o estudo de viabilidade possibilita decisões necessárias sobre o empreendimento no início de qualquer projeto, sendo inconcebível arcar com os custos dos projetos para depois analisar a sua viabilidade da execução. Além de diminuir os riscos financeiros, o plano de negócios ajuda o empreendedor a definir as decisões do negócio com clareza e inteligência antes de aplicá-las no mercado.

A análise de concorrentes como parte do plano de negócios não deixa de ser um ponto que deve ser bem analisado antes da tomada das decisões definitivas. Para Toledo (2015), a concorrência é a solução de muitos problemas e os concorrentes devem compartilhar suas dificuldades. O autor indicou que seguir uma identidade sem espelhar na concorrência é bom, mas quanto mais o empreendedor souber sobre o mercado e os seus concorrentes, mais embasamento ele terá para tomar as decisões e cometerá menos erros. Nesta ideia, a análise de concorrentes se torna ainda mais importante nas tomadas de decisões. Mesmo que a maioria das empresas não deixe aberto os seus problemas para seus concorrentes, é possível tomar conhecimento sobre estes problemas e analisá-los na melhor forma. A recuperação de informação pode ser utilizada para encontrar estas informações.

2.2 RECUPERAÇÃO DE INFORMAÇÃO

A recuperação de informação é um sub-campo da informática que trata do armazenamento automatizado e recuperação de documentos (FRAKES; BAEZA-YATES, 1992). Este conceito se refere a todas as atividades relacionadas à organização, processamento e acesso a informações de todas as formas e formatos (BAEZA-YATES; RIBEIRO-NETO et al., 1999).

Esta área é muito útil em sistemas que procuram produzir informações relevantes para seus usuários, como o Lise.

As informações podem ser buscadas em documentos, bancos de dados ou até na Internet. Quando a informação é buscada na Internet, ela não pode ser solicitada de qualquer forma. Neste caso, geralmente, a recuperação de informação é feita usando combinações de palavras-chave (MOHAMMADIAN, 2004). Segundo Mohammadian (2004), isto requer que os usuários insiram palavras-chave ou seus sinônimos para serem correspondidos com a palavra-chave indexada. O Lise segue esta forma para realizar as pesquisas na Internet ou no banco de dados para recuperar as informações. Após isso, estas informações encontradas passam em uma série de tarefas para gerarem informações mais úteis. Estas tarefas fazem parte da mineração de dados, que é mencionada na próxima seção.

2.3 MINERAÇÃO DE DADOS

A mineração de dados refere-se a extrair o conhecimento de grandes quantidades de dados (HAN; PEI; KAMBER, 2011). Os autores indicaram que nos últimos anos a mineração de dados cresceu muito no setor de informação por causa da disponibilidade de enorme quantidade de dados e da necessidade iminente de transformar estes dados em informações e conhecimentos úteis. Segundo Han, Pei e Kamber (2011), a mineração pode ser usada em várias aplicações como análise de mercado, detecção de fraude, retenção de clientes, controle de produção e exploração científica.

Linoff e Berry (2011) mencionaram seis tarefas da mineração de dados que poderiam resolver muitos problemas de interesse intelectual, econômico e comercial: a classificação, a estimativa, a predição, a regra de associação, o agrupamento e o perfil. A classificação consiste em examinar os recursos de um objeto recém-apresentado e atribuí-lo a um conjunto pré-definido. A estimativa estabelece algum valor a uma variável de acordo com a classificação. A predição funciona da mesma forma, exceto que os registros são classificados de acordo com algum comportamento futuro. A regra de associação determina quais valores deverão estar juntos. O agrupamento é tarefa de segmentar uma população heterogênea em vários subgrupos mais homogêneos. O perfil descreve o padrão encontrado baseado ou não a uma variável. Neste trabalho será útil um pouco destas tarefas para desenvolver o Lise.

2.4 ANÁLISE DE SENTIMENTO

A análise do sentimento é o estudo computacional das opiniões, avaliações, atitudes e emoções das pessoas em relação a entidades, indivíduos, questões, eventos, tópicos e seus atributos (LIU; ZHANG, 2012). Os autores também viram nesta técnica uma boa forma de ajudar as empresas a entender as opiniões dos clientes. Para isso, os *feedbacks* dos clientes são avaliados. A mesma metodologia é usada neste trabalho para descobrir os pontos negativos e

positivos dos concorrentes.

Pang (2008) dividiu a análise de sentimento em duas etapas: etapa de classificação e extração e etapa de sumarização. Na etapa de classificação e extração, o autor considerou duas polaridades: positivo e negativo. Mas nem sempre uma opinião tem somente uma polaridade, a maioria das opiniões tem tanto a polaridade positiva quanto negativa, ou até mesmo a polaridade neutra. Neste último caso é quando não há termos positivos nem negativos. Por isso, o autor avaliou a polaridade de cada termo da opinião e depois juntou eles. Na etapa de sumarização, o autor se preocupou em agregar e representar de forma rápida os principais pontos em uma única revisão, seja de uma ou mais opiniões. Esta metodologia será muito útil no desenvolvimento deste trabalho para resumir quais são os pontos fortes e fracos dos concorrentes segundo os clientes deles.

O desempenho de diversos algoritmos de análise de sentimento para a rede social *Twitter*¹ foi testado por Araújo (2013), como *Emoticons* e *SentiWordNet*. Neste teste foram calculadas as medidas de revocação, precisão, exatidão e *f-measure*. Como a maioria dos algoritmos analisa palavras em inglês, Balahur (2012) investigou uma metodologia de traduzir as palavras com o *Google Translate*² para analisar as expressões de outra língua. Como a metodologia de Balahur foi bem aceita, Reis (2015) utilizou sentenças em diversas línguas, inclusive português, para analisar vários algoritmos, incluindo todos aqueles utilizados no trabalho de Araújo (2013) e outros adicionais, como o *Vader*, o algoritmo utilizado no Lise. No trabalho de Reis (2015) foi observado que as sentenças traduzidas não interferem, de maneira significativa, nos resultados dos algoritmos. Por isso, o Lise irá trabalhar da mesma forma, traduzindo as sentenças para aplicá-las no *Vader*. O *Vader* foi criado para tratar especialmente de texto de mídias sociais e se encontra na biblioteca do *Python* (HUTTO; GILBERT, 2014).

Além de todos esses conceitos é importante conhecer as formas automatizadas atuais que o empreendedor pode utilizar para realizar a análise de concorrentes. Na próxima seção são referenciadas estas ferramentas.

2.5 FERRAMENTAS DE APOIO AOS EMPREENDEDORES

Entre diversos tipos de ferramentas que os empreendedores podem usar para realizar a análise, foram identificados dois tipos: ferramentas de pesquisa com análise automatizada e ferramentas que coletam dados públicos da Internet para realizar a análise automatizada. No primeiro tipo, a *SurveyMonkey*³ é uma destas ferramentas, onde oferecem recursos para o pesquisador criar questionários e distribuí-las. Ela não é especializada em análise de concorrentes e sim para pesquisas em geral. Desta forma, o empreendedor poderá fazer questionário para os clientes sobre a concorrência. Este sistema mostra as respostas analisadas por meios de gráficos

¹ <https://twitter.com/>

² <https://translate.google.com.br/>

³ <https://www.surveymonkey.com/>

e na conta paga oferece outros recursos mais avançados como tabulação cruzada na análise de questionário. Existem várias ferramentas parecidas com estas, como *Google Forms*⁴, *Sua Pesquisa*⁵ e *Survio*⁶.

No segundo tipo identificado, as ferramentas servem para acompanhar a concorrência digital, tendo uma visão ampla do mercado. Algumas delas são *Alexa*⁷, *Open Site Explorer*⁸ e *SEMRush*⁹. Elas utilizam conteúdo da Internet para mostrar informações sobre o assunto pesquisado. A primeira das três mencionadas é a mais avançada, principalmente para acompanhar o mercado internacional. A *Alexa* fornece dados de tráfego de Internet, cria rankings e realiza análise de dados. Porém, a pesquisa deve utilizar dados de concorrentes da região que o empreendedor deseja desenvolver a empresa para ter mais sucesso na análise de concorrentes (GOMES, 2005). Com isso, percebe-se um ponto negativo que este tipo de ferramenta tem para analisar concorrentes de loja física, por causa da grande amplitude do mercado que elas utilizam para pesquisar, ao invés de considerar apenas a localidade que o empreendedor deseja.

RESUMO

Neste capítulo foram descritos os principais conceitos para o desenvolvimento desse projeto. Também foram descritas as principais ferramentas que os empreendedores podem utilizar para realizar a análise de concorrentes. Sobre estas ferramentas foram relatados como os empreendedores poderiam utilizá-las e quais problemas teriam. Depois desse referencial teórico, será mais fácil de entender a metodologia do desenvolvimento desse projeto, que será relatada no próximo capítulo.

⁴ <https://www.google.com/forms/>

⁵ <http://www.suapesquisa.com.br/>

⁶ <http://www.survio.com.br/>

⁷ <http://www.alexa.com/>

⁸ <https://moz.com/researchtools/ose/>

⁹ <https://pt.semrush.com/>

3 METODOLOGIA

A metodologia desse projeto destaca a mineração de dados, que é a principal atividade do Lise. Para isso, algumas tecnologias serão utilizadas e cada uma foi escolhida considerando a funcionalidade desejada e a facilidade de integração no sistema. Para comprovar a eficiência deste método foi realizada uma verificação utilizando a forma manual de analisar os concorrentes e, logo após, comparações de resultados das duas formas: manual e automática.

3.1 RECURSOS TECNOLÓGICOS

O trabalho foi desenvolvido em *Python*, na sua versão 3, uma aplicação *web* com o *framework Django*. Esta plataforma foi escolhida pelo fato de oferecer um desenvolvimento rápido e facilitado. Além disso, a linguagem é multiparadigma e tem ótimos recursos em sua biblioteca, inclusive aqueles necessários para esse sistema. O módulo gerenciamento de dados, que será detalhado no capítulo do Lise, onde ocorre a coleta, a análise, a classificação e a sumarização dos dados, utiliza diversas aplicações e algoritmos. Para seleção destes recursos foram testados outros com funcionalidades parecidas. A escolha baseou-se nos recursos mais adequados para o desenvolvimento do sistema. Cada recurso contribui com pequenas funcionalidades, e juntas gerenciam os dados. Estas ferramentas são:

- **Google Maps API Web Services¹**: Através desta API que o Lise coleta os dados dos estabelecimentos do *Google Maps*. Neste trabalho será utilizado o plano padrão, que é gratuito e público, pois os recursos deste plano são suficientes para utilização no Lise.
- **Google Cloud Translation API²**: Esta API será utilizada na tradução de sentenças que serão analisadas pela ferramenta de análise de sentimento. Este tradutor já foi testado no trabalho de Balahur (2012). Diferente da ferramenta anterior, ela não é gratuita. Os preços da *Translation API* baseiam no uso, cobrando por quantidade de caracteres traduzidas. No Apêndice A mostra porque o *Google Translation* foi escolhido para realizar o serviço de tradução neste trabalho.
- **Facebook Graph API³**: Este recurso se trata da API da rede social *Facebook*. Nesse trabalho, ele é utilizado somente para relacionar os concorrentes com as suas *fanpages*. Porém nenhuma informação delas é coletada, apenas é registrado que o estabelecimento possui ou não *fanpage*.
- **CoGrOO**: O Corretor Gramatical para *OpenOffice* visa verificar os erros gramaticais como o acordo nominal e verbal, crase, regência nominal e verbal, mau uso do

¹ <https://developers.google.com/maps/>

² <https://cloud.google.com/translate/>

³ <https://developers.facebook.com/docs/graph-api/>

adjetivo "mau"(mau) e do advérbio "mal"(mal), entre outros erros comuns que podem ser encontrados no português brasileiro (KINOSHITA; SALVADOR; MENEZES, 2006). Além destas utilidades, neste trabalho esta aplicação será útil com suas funcionalidades na análise textual, como *stemmers*, *tokenization* e *taggers*. *Stemmers* é o processo que reduz as palavras flexionadas à sua raiz. *Tokenization* é o processo de demarcação, classificando seções de uma sequência de caracteres de entrada. Por fim, os *taggers* criam marcadores morfossintáticos das palavras.

- **Vader:** O método escolhido para fazer a análise de sentimento foi o *Vader* (HUTTO; GILBERT, 2014). Ele está entre os métodos testados no trabalho de Reis (2015).
- **Django Q e Redis:** No Lise, estes dois recursos trabalham sempre juntos para gerenciar tarefas recorrentes. Isto é necessário para programar o momento e a frequência que o sistema irá minerar os dados. O *Django Q* é um aplicativo de fila, agendador e trabalhador nativo do *Django* usando o multiprocessamento *Python* (Q, 2018b). Mas esta aplicação precisa de um corretor de mensagem para ficar entre ele e o *framework Django*, gerenciando os pacotes de tarefas. O *Redis* foi o corretor escolhido por ser simples e fácil, além de ser o corretor padrão do *Django Q* (Q, 2018a).

3.2 BASE DE DADOS

Como foi dito no capítulo anterior, a análise de concorrentes precisa de dados reais sobre o mercado dentro da região desejada de empreendimento. Para realizar a análise de concorrentes de forma automatizada, será preciso ter a coleta de dados também automatizada. A API do *Google Maps* oferece dados suficientes sobre empresas para fazer esta análise. Por isso, ela será utilizada, frequentemente, para coletar os dados e, desta forma, manter a análise atualizada com o mercado.

A API do *Google Maps* limita até sessenta localidades por pesquisa e cinco avaliações por cada localidade, seja no plano gratuito ou pago. Então, por uma categoria de negócio que tem apenas uma palavra-chave, na primeira coleta irão ser coletados no máximo 300 avaliações. Observando o sistema do *Google Maps*, foi percebido que a maioria das avaliações contém mais de uma opinião. Para aumentar ainda mais a amostra de dados na análise, cada avaliação será separada se houver mais de uma opinião. Este procedimento irá ser feito por algoritmo criado nesse projeto, composto por vários componentes auxiliares mencionados na seção anterior.

Com esta base de dados, será possível sumarizar as informações e gerar uma das principais informações sobre análise de concorrentes: os pontos fortes e fracos da concorrência de uma categoria de negócio. Além da sumarização, isto vai depender da classificação de opinião entre positivo, negativo e neutro. Para resolver estas questões, foram utilizados componentes como CoGrOO e *Vader*. Este processo é explicado, detalhadamente, no capítulo do Lise.

Para manter a base de dados atualizada, o sistema irá fazer *auto update* frequentemente, utilizando o *Django Q* e *Redis*. Esta atualização passará em cada categoria cadastrada no sistema e fará todo o procedimento de coleta e análise de dados, adicionando apenas aqueles dados que não estão no banco de dados.

3.3 VERIFICAÇÃO DO SISTEMA

Mesmo sabendo que já foram utilizadas metodologias parecidas com a deste trabalho em Xu (2011), é necessário verificar se os resultados produzidos no Lise são verídicos e consideráveis. Para isso, foram analisadas as seguintes hipóteses:

H_0 - O Lise tem uma proporção de acertos maior ou igual a 75% sobre as análises das avaliações dos concorrentes.

H_a - O Lise tem uma proporção de acertos menor que 75% sobre as análises das avaliações dos concorrentes.

Uma amostra é baseada na quantidade da população. Segundo (FILHO, 2009), população infinita é denominada a qualquer população cuja quantidade de elementos seja muito grande ou seja difícil de ser quantificada. É o caso da população de avaliações do Google Map, pois existe um número indefinido de avaliações que cresce a todo momento. Miot (2011) apresenta a fórmula para calcular a amostra de variáveis qualitativas em uma população infinita:

$$n = \left(\frac{Z_{\frac{\alpha}{2}} \times \sqrt{p \times q}}{E} \right)^2$$

Equação 1

Onde:

- n é o tamanho da amostra;
- $Z_{\frac{\alpha}{2}}$ representa o valor crítico para o grau de confiança desejado;
- p é a proporção de resultados favoráveis da variável na população;
- q é a proporção de resultados desfavoráveis na população ($q = 1 - p$);
- E representa o erro padrão desejado.

O grau de confiança desejado é de 95%, por isso o valor crítico foi de 1,96, seguindo a tabela de distribuição normal padrão (REIS et al., 1999), que está no Anexo A. As proporções de resultados favoráveis e desfavoráveis não podem ser definidas, pois o possível resultado desta pesquisa é desconhecido. Neste caso é recomendado utilizar o valor de 50% para cada proporção (OCHOA, 2013). Por fim, o valor do erro padrão utilizado foi 5%. Com base nesses valores o

tamanho da amostra foi calculada aplicando a fórmula de Miot (2011) mostrada na Equação 1, encontrando o valor aproximado de 384 avaliações:

$$n = \left(\frac{1.96 \times \sqrt{0.5 \times 0.5}}{0.05} \right)^2 \Rightarrow n = \left(\frac{0.98}{0.05} \right)^2 \Rightarrow n = (19.6)^2 \Rightarrow n \cong 384$$

Nas avaliações há duas definições para serem analisadas: polaridade e sumarização. A polaridade consiste em dizer se as avaliações são positivas, negativas ou neutras. A sumarização categoriza as avaliações em algum tema empresarial, como a qualidade. Mas, para realizar estas duas análises foi criada uma base de dados com quantidade maior que 384. É preferível que esta amostra seja dividida em grupos com categorias e locais diferentes. Isto porque o Lise poderá analisar os concorrentes em diversas categorias empresariais e em diferentes locais do Brasil.

Com a base de dados da amostra, a polarização e a sumarização foram analisadas de forma manual e, em seguida, de forma automática. Assim que as duas análises foram concluídas, foram comparados os resultados obtidos e verificados as proporções de acertos sobre a amostra. Com esta proporção é possível realizar os testes de hipóteses baseados em dados contados. Freund (2009) apresenta a fórmula para calcular o valor crítico para testes relativos a proporções de grandes amostras:

$$z = \frac{x - n \times p_0}{\sqrt{n \times p_0 \times (1 - p_0)}}$$

Equação 2

Onde:

- z representa o valor crítico;
- x é a proporção dos dados contados;
- n é a amostra total;
- p_0 é a probabilidade da hipótese.

Após encontrar o valor crítico, as hipóteses são analisadas de acordo com os critérios da Tabela 1, apresentada por Freund (2009). Segundo o autor, na tabela o z_α representa o valor crítico do nível de significância. O nível de significância é a diferença entre 1 e o nível de confiança. Como o nível de confiança definido para a amostra foi de 95%, então o nível de significância será de 5%. Com isso, o valor crítico do nível de significância é de 1,65, seguindo a tabela de distribuição normal (REIS et al., 1999), que está no Anexo A. De acordo com a hipótese alternativa definida no início dessa seção, sua expressão formada é $p < 0,75$, a primeira linha da Tabela 1. Por tanto, para a hipótese nula ser aceita, o z calculado na fórmula de Freund (2009) deve ser maior que -1,65.

Tabela 1 – Critérios para teste de hipótese

Hipótese alternativa	Rejeitar a hipótese nula se	Aceitar a hipótese nula se
$p < p_0$	$z \leq -z_\alpha$	$z > -z_\alpha$
$p > p_0$	$z \geq z_\alpha$	$z < z_\alpha$
$p \neq p_0$	$z \leq -z_{\frac{\alpha}{2}}$ ou $z \geq z_{\frac{\alpha}{2}}$	$-z_{\frac{\alpha}{2}} < z < z_{\frac{\alpha}{2}}$

Fonte: Elaborada pela autora baseada em Freund (2009)

RESUMO

Neste capítulo foi descrita a criação e manutenção da base de dados. Mas antes disto, para entender as principais tecnologias usadas nesta atividade, elas foram conceituadas em recursos tecnológicos. Por fim, houve uma descrição de como esta metodologia será validada para comprovar sua utilidade na análise de concorrentes.

4 LISE

O projeto Lise é uma aplicação *web* com funcionalidades para qualquer tipo de empreendedor realizar a análise de concorrentes. O nome surgiu do final da palavra “análise”. No desenvolvimento foram definidas algumas nomenclaturas para padronizar os objetos. Compreender estas nomenclaturas facilitará o acompanhamento do que foi feito no projeto. Além disso, a lista de requisitos e os diagramas podem esclarecer as funcionalidades do projeto Lise.

4.1 NOMENCLATURAS DOS OBJETOS

Para melhor entendimento das nomenclaturas definidas no projeto, a Figura 1 mostra uma estrutura em relação ao grau de hierarquia dos objetos. Desta forma, para cada categoria de negócio pode haver vários concorrentes, para cada concorrente pode haver várias avaliações, e assim por diante.

Figura 1 – Hierarquia dos objetos



Fonte: Elaborada pela autora

No topo da pirâmide estão as categorias, que tratam do tipo de negócio especificado pelo usuário. Abaixo disso, os concorrentes são todos os estabelecimentos que tenham a mesma categoria de negócio. As avaliações são comentários dados pelos clientes do estabelecimento.

As opiniões são sentenças mais curtas sobre algum assunto empresarial. Por fim, o tópico é a palavra-chave da opinião, ou seja, sobre qual campo empresarial a sentença trata. Por exemplo, na opinião “Atendimento não agradou” o tópico deverá ser atendimento. Estes tópicos são os pontos dos concorrentes que o empreendedor deve saber para criar estratégias de mercado.

4.2 REQUISITOS

O levantamento de requisitos foi realizado com base nos objetivos desse projeto e nas necessidades para análise de concorrentes, segundo as indicações do Sebrae. Na Tabela 2 está a lista dos principais requisitos do Lise separados com o código, a descrição e o tipo. O código do requisito servirá para mencionar o requisito com mais facilidade nas especificações de casos de uso no Apêndice B. A descrição se refere ao requisito em si. O tipo de requisito indica se o requisito é funcional ou não funcional.

Tabela 2 – Requisitos funcionais e não funcionais do Lise

Código	Descrição	Tipo
R1	O sistema deverá analisar os concorrentes dos planos de negócios de qualquer estado brasileiro.	Funcional
R2	O usuário poderá cadastrar seus negócios planejados.	Funcional
R3	O usuário poderá solicitar novas categorias de negócios.	Funcional
R4	A base de dados deverá ser frequentemente atualizada com novos dados dos negócios cadastrados.	Não Funcional
R5	O sistema deverá ser desenvolvido para plataforma <i>web</i> .	Não Funcional
R6	O sistema deverá coletar os dados do <i>Google Maps</i> .	Funcional
R7	O administrador do Lise poderá cadastrar, deletar, atualizar e visualizar as categorias de negócios.	Funcional
R8	O administrador do Lise deverá relacionar as palavras aos tópicos e às categorias de negócio, como também pode recusá-las caso não pertença a nenhum tópico.	Funcional
R9	O administrador e o empreendedor poderão cadastrar, editar, entrar e sair na sua conta.	Funcional
R10	O sistema deverá enviar notificação automaticamente para o usuário ao cadastrar um negócio ou solicitar uma categoria. O administrador poderá enviar uma notificação quando quiser.	Funcional
R11	O administrador poderá visualizar as atualizações de dados para ter o controle e o histórico do que está atualizando.	Funcional

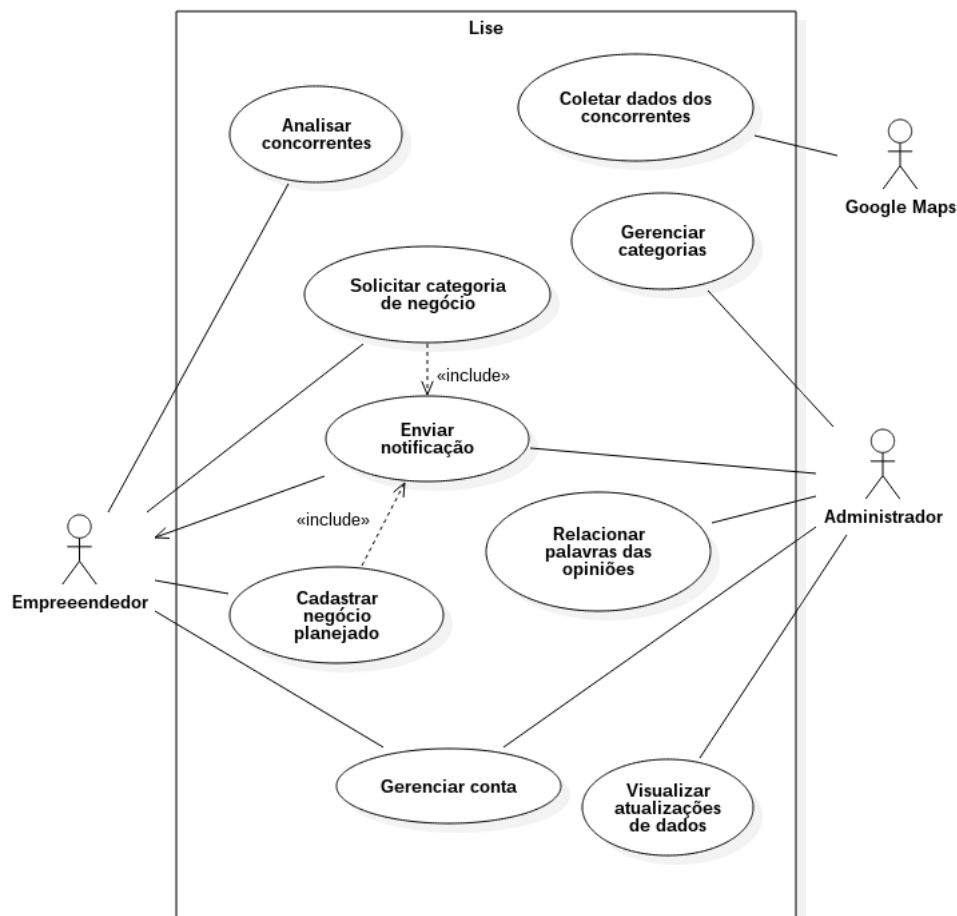
Fonte: Elaborada pela autora

4.3 DIAGRAMAS

4.3.1 Casos de Uso

Os casos de uso foram baseados nos requisitos funcionais mencionados na seção anterior. Os mais importantes do sistema estão sendo representados na Figura 2. As especificações dos principais casos de uso deste diagrama estão no Apêndice B deste documento. Estas especificações seguiram o modelo ensinado por Wazlawick (2010).

Figura 2 – Diagrama de casos de uso

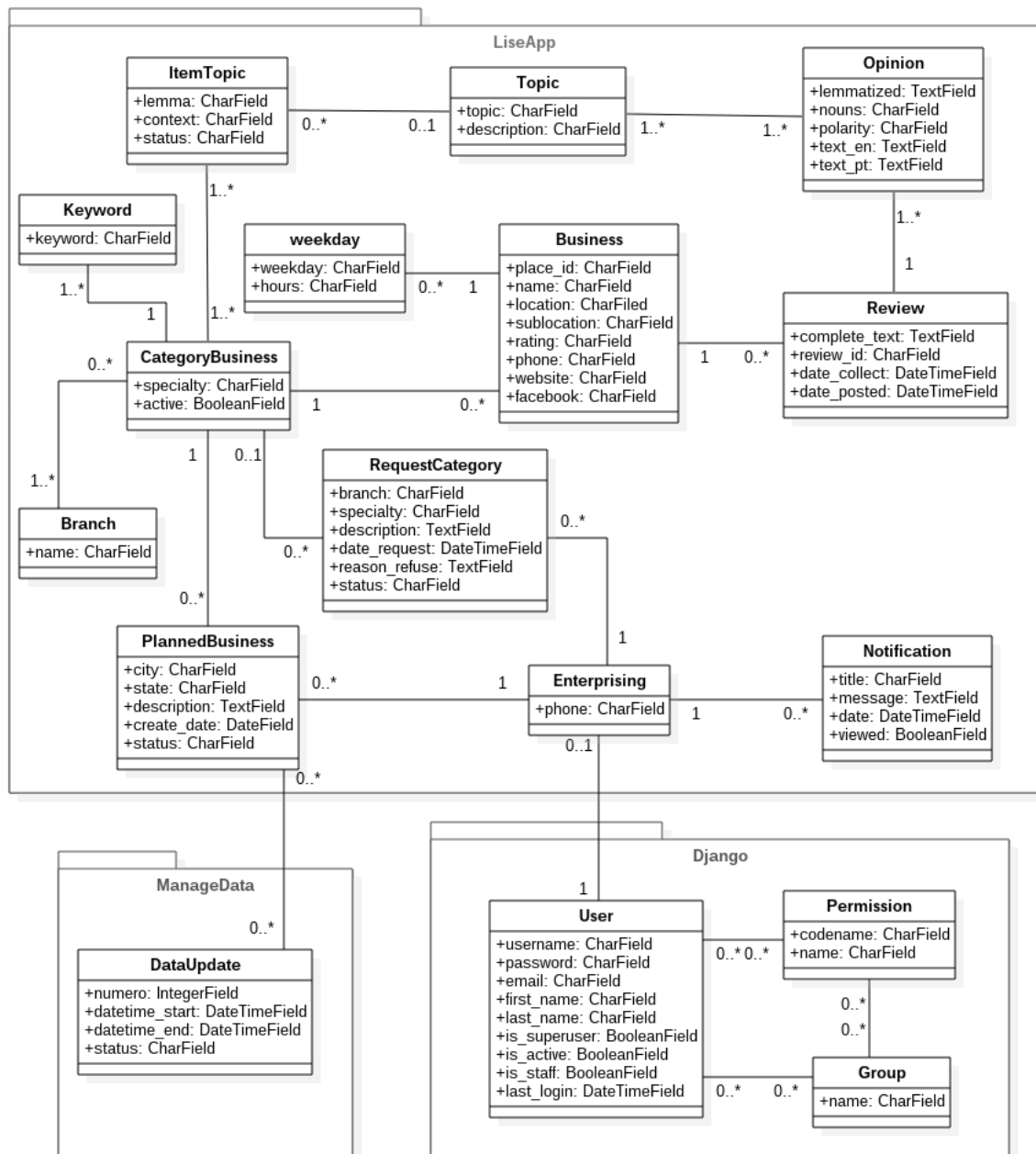


Fonte: Elaborada pela autora

4.3.2 Diagrama de Classes

O diagrama UML de classes do Lise é apresentado na Figura 3, onde estão as principais classes relacionadas. Além das classes criadas durante o desenvolvimento do projeto, a Figura 3 também mostra as principais classes padrões do *Django* utilizadas.

Figura 3 – Diagrama UML de classes



Fonte: Elaborada pela autora

A descrição destas classes estão na lista abaixo:

- **User:** A classe do usuário é padrão do *Django*. Nesta classe estão os atributos e os métodos necessários para autenticação no sistema.
- **Permission:** Também é uma classe padrão do *Django*. Através desta classe é possível gerenciar permissões de usuários e grupos.
- **Group:** A classe *Group* é outra das principais classes padrões do *Django*. Ela relaciona os usuários que necessitam de condições iguais. Por exemplo, os administradores

devem ter as mesmas permissões.

- **Enterprising:** São os empreendedores cadastrados no Lise. Para poder analisar os concorrentes, o usuário deve ser desta classe, o que não acontece com os administradores, que são usuários mas não empreendedores.
- **Notification:** Representa as notificações dos empreendedores. Informações como a mensagem de uma notificação ou a data da notificação estão nesta classe.
- **PlannedBusiness:** São os negócios planejados que os empreendedores desejam analisar os concorrentes. Estes empreendedores podem ter mais de um negócio para analisar ao mesmo tempo.
- **Branch:** Trata dos ramos das categorias, que agrupam uma lista de especialidades, que são as categorias. Porém, cada categoria pode fazer parte de mais de um ramo. Por exemplo, no ramo alimentício pode ter as especialidades pizzaria, restaurante ou sorveteria. Com esta separação será mais fácil para o empreendedor definir o negócio que deseja analisar.
- **CategoryBusiness:** Esta classe representa as categorias dos negócios e podem ser gerenciadas pelo administrador do Lise.
- **Keyword:** Representa as palavras-chave das categorias. São estas palavras utilizadas para pesquisar concorrentes junto com a localidade desejada. Uma categoria pode ter várias palavras e de cada palavra será feita uma pesquisa. A lista de concorrentes é definida pela união dos resultados de todas as pesquisas.
- **RequestCategory:** Nesta classe ficam as instâncias das solicitações de novas categorias realizadas pelo empreendedor.
- **Business:** São os concorrentes envolvidos nas análises de concorrentes. As instâncias desta classe contêm as informações das empresas.
- **Weekday:** Os dias de funcionamento das empresas são registrados nesta classe. Cada instância contém o dia da semana, o horário de funcionamento e a empresa tratada. Mas, algumas empresas não disponibilizam esta informação. Por isso, alguns concorrentes não têm relações com esta classe.
- **Review:** São as avaliações de uma empresa. Estas avaliações são as originais que foram coletadas do *Google Maps*.
- **Opinion:** É cada sentença da avaliação considerada como opinião. Isso porque cada avaliação poderá ter mais de uma opinião.
- **Topic:** Representa um assunto empresarial que a opinião trata. Uma opinião também pode ter mais de um tópico. Também um tópico pode tratar mais de uma opinião.
- **ItemTopic:** São palavras relacionadas com algum tópico. É através destas palavras que o Lise identifica sobre o que uma avaliação trata.

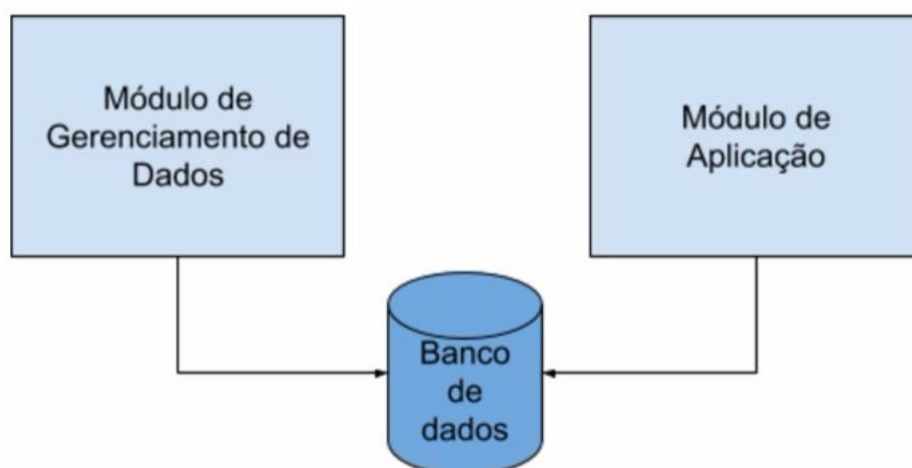
- **DataUpdate:** Nesta classe ficam todas as instâncias das atualizações de dados. Com isso, o administrador consegue manter o controle e o histórico das atualizações realizadas.

4.4 ARQUITETURA

De acordo com a classificação de Sommerville (2011), o padrão de arquitetura utilizado é de repositório. Segundo o autor, neste padrão todos os dados em um sistema são gerenciados em um repositório central, acessível a todos os componentes do sistema. Os componentes não interagem diretamente, apenas por meio do repositório. No caso do Lise, o banco de dados é o repositório compartilhado entre dois módulos, que só se interagem através do banco de dados. O Lise foi arquitetado desta maneira para agir com mais rapidez. Além disso, com o tempo o banco de dados irá armazenar uma grande quantidade de registros que irão continuar sendo consultados frequentemente.

Dessa forma, a arquitetura do Lise é dividida em dois módulos: gerenciamento de dados e aplicação. O primeiro é encarregado de movimentar o banco de dados. Todas as etapas de mineração de dados são realizadas neste módulo. No segundo foram desenvolvidas as funcionalidades em geral do Lise. A Figura 4 representa esta arquitetura. Desta forma, uma funcionalidade não realiza o processo de mineração diretamente, apenas consulta a informação no banco de dados, que anteriormente foi preenchida pelo módulo de gerenciamento de dados. Este primeiro módulo deve ser frequentemente executado para atualizar o banco de dados e assim o módulo de aplicação poderá mostrar informações reais de acordo com o mercado.

Figura 4 – Arquitetura do Lise

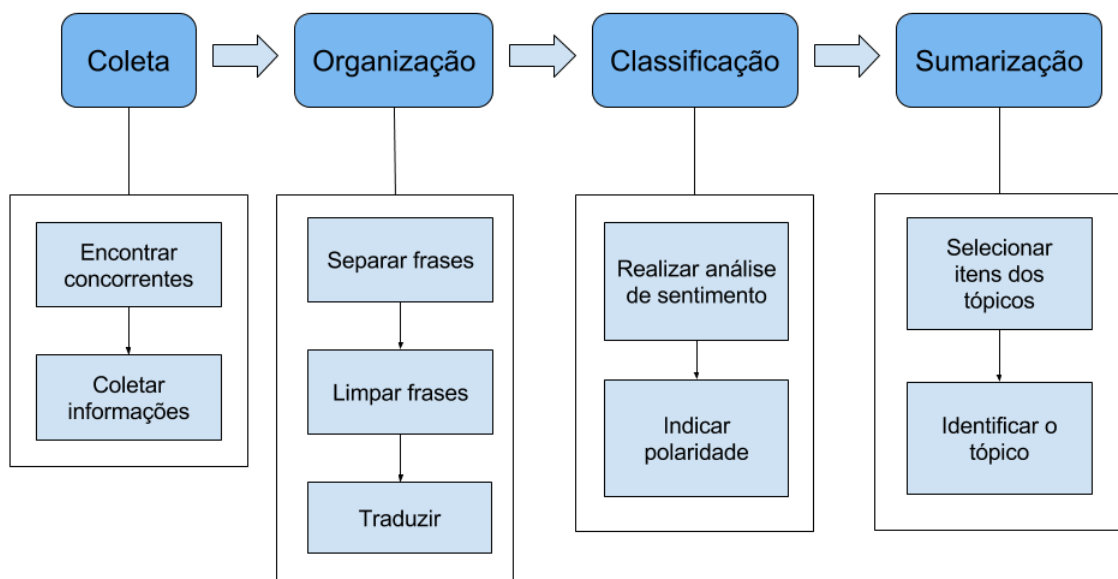


Fonte: Elaborada pela autora

4.4.1 Módulo de Gerenciamento de Dados

Como já foi dito anteriormente, o módulo de gerenciamento de dados trata da mineração dos dados do *Google Maps*. A Figura 5 representa a arquitetura que foi seguida para o desenvolvimento deste módulo. O processo será realizado em sequência, pois o resultado de uma fase se constitui na entrada da outra. Em resumo, os dados foram coletados, organizados, classificados e sumarizados.

Figura 5 – Arquitetura do módulo de gerenciamento de dados



Fonte: Elaborada pela autora

- **Coleta:** Esta fase tem o objetivo de coletar as informações dos estabelecimentos concorrentes como avaliações, nomes, endereços, horários de funcionamento, entre outros. Mas antes estes estabelecimentos concorrentes são encontrados através de pesquisas com as palavras-chave da categoria do negócio planejado que foi cadastrado. Este procedimento será realizado através da API do *Google Maps*.
- **Organização:** A organização pode ser vista como uma preparação dos dados da fase anterior para a fase posterior. Isto porque os dados recolhidos na coleta não estão adequados para os algoritmos de análise de sentimentos. A maioria destes algoritmos analisa somente palavras em inglês. Além disso, algumas mensagens têm erros sintáticos e mais de um ponto de vista do cliente. Por isso, é necessário a separação e a limpeza de dados. Nesta fase será utilizada a API do *Google Translate* para realizar a tradução e a ferramenta CoGrOO para tratar o texto das frases.
- **Classificação:** Na classificação, as sentenças tratadas na fase anterior são analisadas e logo depois, para cada sentença, de acordo com o valor da polaridade, é indicado

uma dessas polaridades: negativa, positiva ou neutra. Como descrito no capítulo de metodologia, a ferramenta de análise de sentimento utilizada é o *Vader*.

- **Sumarização:** Nesta última fase do processo de gerenciamento de dados, acontece a identificação do tópico que a sentença trata. Isto acontece através da relação do tópico com os itens dos tópicos que foi realizado pelo administrador. Para isso, é novamente necessário o serviço CoGrOO. As novas palavras não cadastradas são selecionadas pelo Lise para o administrador relacioná-las com os tópicos.

Estas fases são executadas com base nas especificações do cadastro do negócio planejado. Para esclarecer ainda mais as fases deste processo, a Tabela 3 exemplifica os resultados de cada fase. Na demonstração foi citado apenas um exemplo de cada campo analisado.

Tabela 3 – Exemplo de resultados das fases de gerenciamento de dados

Fase	Subfase	Resultado
Coleta	Encontrar concorrente	Concorrente X
	Coletar informações	"a comida é muito boa mas o atendimento nao me agradou :/"
Organização	Separar frases	1. "a comida é boa" 2. "atendimento não me agradou :/"
	Limpar frases	Comida é boa
	Traduzir	<i>Food is good</i>
Classificação	Realizar análise de sentimento	Positivo: 0.395 Negativo: 0.0 Neutro: 0.605
	Indicar polaridade	Positivo
Sumarização	Selecionar os itens dos tópicos	comida
	Identificar tópico	Qualidade

Fonte: Elaborada pela autora

4.4.2 Módulo de Aplicação

Neste módulo as funcionalidades do sistema foram tratadas de acordo com os requisitos descritos na seção 4.2. A arquitetura deste módulo segue a mesma do Django, um padrão Model Template View (LAUBE, 2017). Como dito anteriormente, o módulo de aplicação terá ligação com banco de dados, da mesma forma no módulo de gerenciamento de dados, e esses módulos se relacionam através do banco. Portanto, este módulo não minera os dados sobre os concorrentes, apenas realiza consultas nos registros atualizados do banco de dados, para poder organizá-las e disponibilizá-las aos empreendedores.

RESUMO

Neste capítulo houve um esclarecimento das nomenclaturas dos conceitos utilizados neste projeto. Depois foram citados os requisitos funcionais e não funcionais. Em seguida, foram

mostrados o diagrama de caso de uso, que complementou os requisitos, e o diagrama de classes. Por fim, foi feita uma explicação da arquitetura de funcionamento da aplicação, separando por módulo de gerenciamento de dados e módulo de aplicação.

5 RESULTADOS

Os primeiros resultados a serem mostrados são sobre a verificação, que foi descrita no capítulo de metodologia. O resultado da verificação apresenta os valores de acertos e as conclusões dos testes de hipótese da polaridade e da sumarização. Depois disto, são apresentadas as principais interfaces da aplicação para o empreendedor.

5.1 VERIFICAÇÃO

Como foi dito no capítulo de metodologia, foram analisadas as polaridades e as sumarizações das avaliações de forma manual e automática. Os resultados serão apresentados através da proporção de acertos sobre a amostra. Para isso, foi coletada uma amostra de 400 avaliações. Desta amostra foi construída uma base de dados com avaliações de três categorias, cada uma em um local diferente. Estas categorias estão sendo mostradas na Tabela 4.

Tabela 4 – Categorias e localidades utilizadas na verificação

Categoria	Localidade	Quantidade de avaliações
Clínica e diagnóstico de imagem	Teresina, PI	150
Vestuário	Santos, SP	125
Pizzaria	Manaus, AM	125

Fonte: Elaborada pela autora

Na polaridade o valor que poderia ser dado para avaliação está entre positivo, negativo ou neutro. Para a sumarização foi disponibilizada, como base, uma tabela com os tópicos e suas descrições, como mostra na Tabela 5. Nesta tabela estão os principais tópicos empresariais. Com a lista de avaliações, um voluntário analisou cada uma delas e deu sua opinião sobre qual polaridade e quais os tópicos da tabela de sumarização pertenciam. A resposta da polaridade deveria ser entre os três descritos acima, não podendo ficar em branco e nem tendo mais de uma polaridade. A resposta da sumarização deveria ser um ou mais tópicos da Tabela 5, podendo deixar em branco se a avaliação não pertencesse a nenhum dos tópicos da tabela. Com as respostas da análise manual, as respostas do Lise foram geradas automatizadas. Com as duas análises realizadas, foi possível calcular o resultado da polaridade e da sumarização.

Tabela 5 – Tópicos empresarias para basear na sumarização

Tópico	Descrição
Atendimento	Este tópico está relacionado ao tratamento dos funcionários dos concorrentes.
Qualidade	Este tópico trata da essência do produto do concorrente.
Localização	Neste tópico é avaliado o ponto de localidade do estabelecimento.
Preço	Este tópico representa o custo dos produtos ou serviços.
Variedade	Este tópico trata da quantidade de opções disponíveis para os clientes.
Tempo	Neste tópico é avaliado a disponibilidade do tempo de funcionamento do estabelecimento.
Ambiente Interno	Este tópico trata da estrutura ou decoração, como também sobre clima e espaço.
Funcionamento	Neste tópico é avaliado os pontos importantes de como a empresa trabalha para atender clientes.
Geral	Este tópico trata do estabelecimento em geral.

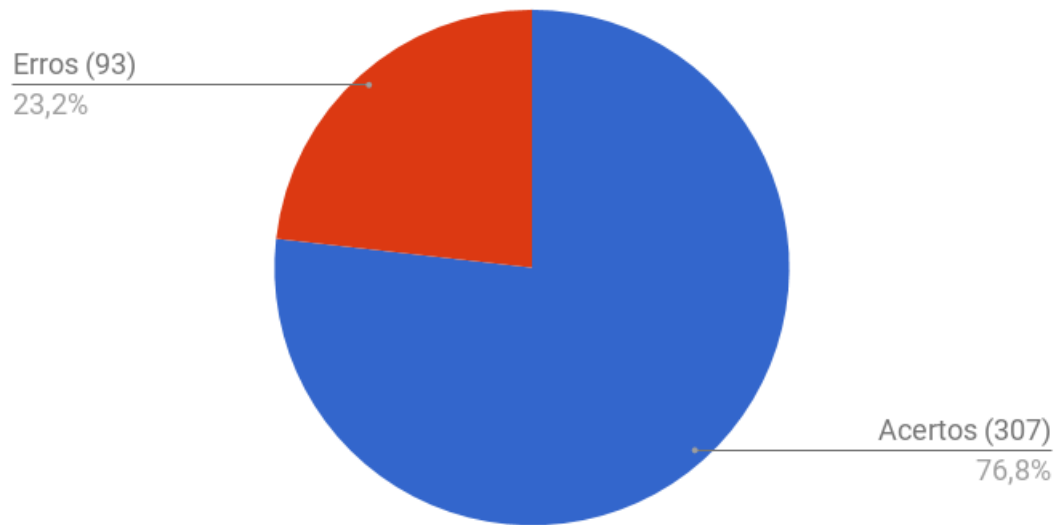
Fonte: Elaborada pela autora

A Figura 6 mostra a proporção de acertos da polaridade. Na polarização, o Lise acertou 307 polaridades das 400 avaliações, conseguindo uma porcentagem de 76,8% de acerto. Utilizando a fórmula do Freund (2009) mostrada na Equação 2, o valor crítico calculado para a polarização foi de, aproximadamente, 0,81:

$$z = \frac{307 - 400 \times 0.75}{\sqrt{400 \times 0.75 \times (1 - 0.75)}} \Rightarrow z = \frac{7}{\sqrt{75}} \Rightarrow z = \frac{7}{8.66} \Rightarrow z \cong 0.81$$

Os critérios para o teste de hipóteses apresentados na Tabela 1 no capítulo de metodologia, mostrou a seguinte conclusão: para aceitar a hipótese nula apresentada no capítulo de Metodologia o valor crítico da hipótese deve ser maior que -1,65. Portanto, a hipótese nula é aceita na polaridade, pois 0,81 é maior que -1,65.

Figura 6 – Resultado da polarização da verificação



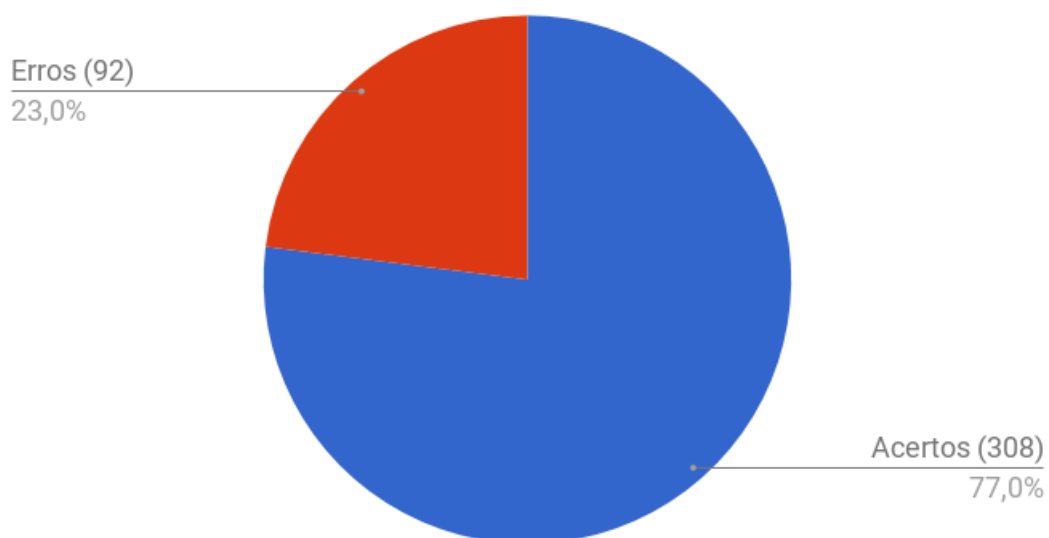
Fonte: Elaborada pela autora

Da mesma forma, a Figura 7 mostra o resultado da sumarização. Nesta análise, o Lise acertou 308 tópicos nas 400 avaliações, conseguindo 77% de acerto. O valor crítico calculado segundo a fórmula Freund (2009) mostrada na Equação 2 foi de, aproximadamente, 0,92:

$$z = \frac{308 - 400 \times 0.75}{\sqrt{400 \times 0.75 \times (1 - 0.75)}} \Rightarrow z = \frac{8}{\sqrt{75}} \Rightarrow z = \frac{8}{8.66} \Rightarrow z \cong 0.92$$

Portanto, a hipótese nula também foi aceita na sumarização, pois 0,92 é maior que -1,65.

Figura 7 – Resultado da sumarização da verificação



Fonte: Elaborada pela autora

5.2 INTERFACE DA APLICAÇÃO

A interface do sistema é baseada no modelo de código aberto *AdminBSBMaterialDesign*¹ e utiliza o *framework Bootstrap* e a linguagem de design *Google Material Design*. Os gráficos são gerados através das bibliotecas de *JavaScript: Chartjs, Flot Chart e Morris*. Com eles é possível criar facilmente recursos interativos para *web*. Além disso, os mapas são gerados pela API do *Google Maps*. Com estes recursos foi construída uma interface atraente e interativa.

Dentre as funcionalidades requeridas no Lise, a principal é produzir informações sobre concorrentes e realizar análises para o empreendedor. Mas para isso, as categorias devem estar cadastradas. Inicialmente foram cadastradas 60 categorias empresariais que o Lise pode analisar. A lista destas categorias está no Apêndice C. Nesta lista, além dos nomes das categorias estão os ramos e as palavras-chaves. Esta lista foi baseada na lista de categorias de empresas que tiveram mais estabelecimentos abertos em 2014 disponível no DataSebrae². Com estas categorias, os empreendedores poderão cadastrar seu plano de negócio ou solicitar uma nova categoria se ela ainda não foi cadastrada. A Figura 8 mostra a página de cadastro de plano de negócios. Depois do cadastro o Lise produzirá as informações e o plano de negócios cadastrado será mostrado na lista de planos de negócios na conta do empreendedor, como mostra na Figura 9.

Figura 8 – Página de cadastro do plano de negócio do Lise

A interface de cadastro de plano de negócio do Lise apresenta um formulário com o título "CADASTRO DE NEGÓCIO". O formulário é dividido em duas colunas principais: "Categoria do Negócio" e "Localidade".

- Categoria do Negócio:**
 - Ramo:** Um menu suspenso com o texto "Selecione um ramo empresarial".
 - Especialidade:** Um menu suspenso com o texto "Selecione uma categoria".
- Localidade:**
 - Estado:** Um menu suspenso com o texto "Selecione um estado".
 - Cidade:** Um menu suspenso com o texto "Selecione uma cidade".

Abaixo das colunas, há um link: "Não tem categoria para sua ideia? Solicite uma categoria.".

Na seção "Descrição", há um campo de texto com o placeholder: "Descreva um pouco sobre o negócio que você planeja. :)".

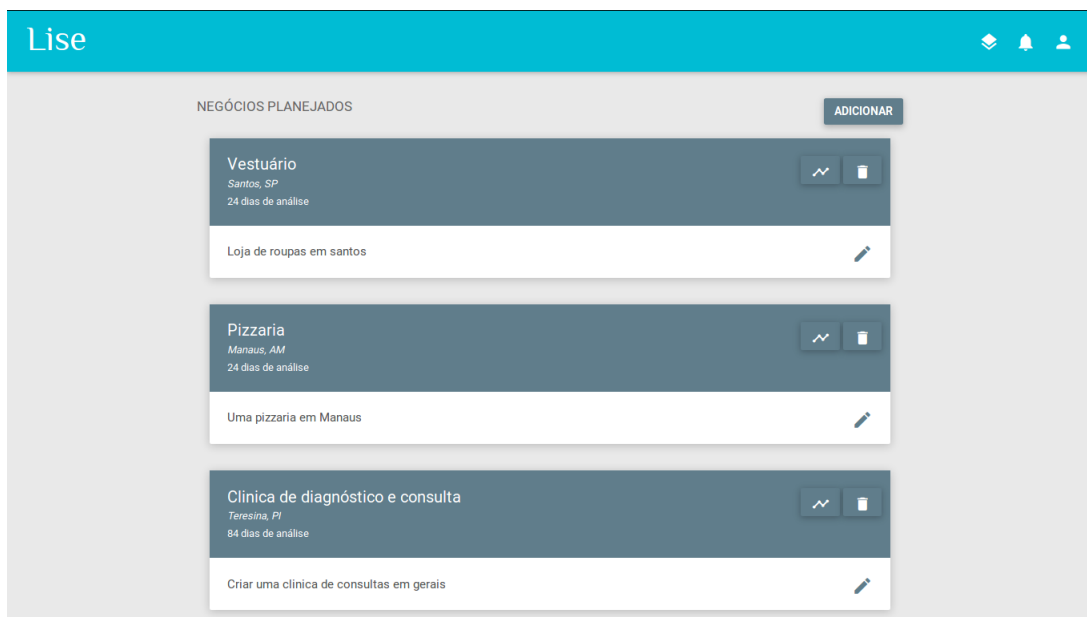
No canto inferior direito do formulário, há dois botões: "Cancelar" e "Cadastrar".

Fonte: Elaborada pela autora

¹ <https://github.com/gurayyarar/AdminBSBMaterialDesign>

² http://sistema.datasebrae.com.br/sites/novo_datasebrae/#Empresas_Total-de-empresas_Tabela

Figura 9 – Lista de planos de negócios do empreendedor do Lise



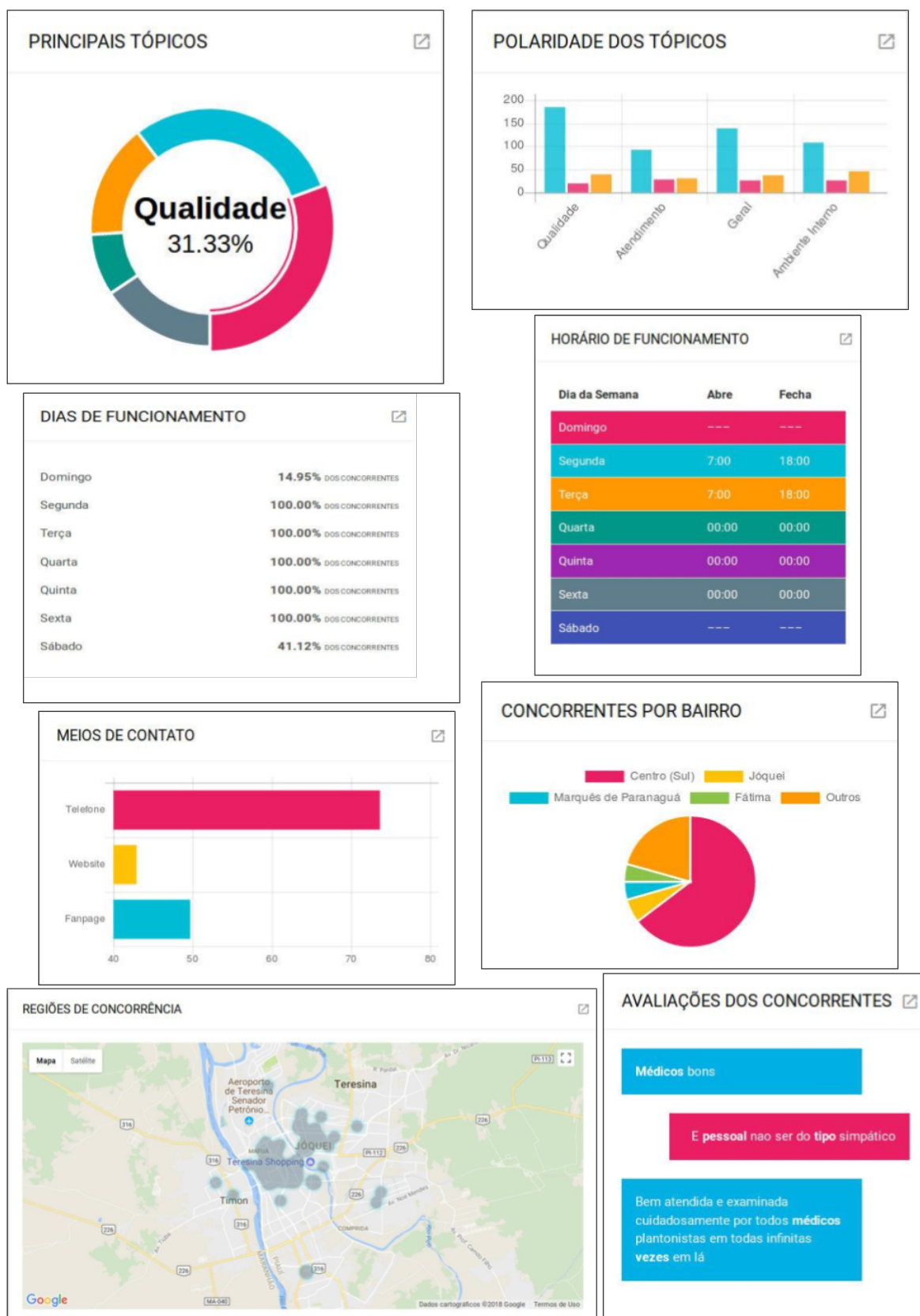
Fonte: Elaborada pela autora

Para informar aos empreendedores sobre os concorrentes facilmente e de maneira simples, a aplicação produzirá gráficos, mapas, tabelas e listas com as informações analisadas de acordo com os dados coletados. A Figura 10 mostra os *cards* das informações apresentados no sistema. As descrições destas informações são:

- **Principais tópicos:** Este *card* apresenta os tópicos mais comentados pelos clientes dos concorrentes.
- **Polaridade dos tópicos:** As porcentagens de avaliações positivas, negativas e neutras de cada tópico são mostradas neste *card*.
- **Dias de funcionamento:** Este *card* mostra os dias da semana que os estabelecimentos funcionam.
- **Horários de funcionamento:** Neste *card* são apresentados os horários que os estabelecimentos estão abertos.
- **Meios de contato:** Neste *card* é mostrada a quantidade de concorrentes que disponibilizaram os meios de contato para o cliente se informar. Os meios de contatos analisados foram: telefone, *website* e *fanpage* do Facebook.
- **Avaliações sobre concorrentes:** Neste *card* são listadas as opiniões positivas e negativas que os próprios clientes dos concorrentes avaliaram.
- **Concorrentes por bairro:** Este *card* apresenta os bairros onde os concorrentes estão na cidade que o empreendedor cadastrou seu negócio.

- **Regiões de concorrência:** Neste *card* as regiões de concentração dos concorrentes são mostradas na localidade que o empreendedor cadastrou seu negócio.

Figura 10 – Modelos dos tipos de informações produzidas



Fonte: Elaborada pela autora

Além deste sistema para os empreendedores analisarem os concorrentes, existe o sistema para administrar estas análises. Na área de *admin* do Lise é possível controlar as análises de concorrentes, as categorias e as solicitações de categorias, as atualizações de dados, tarefas agendadas, entre diversas outras tarefas para administrar o Lise. As Figuras 11, 12, 13, 14, 15 e 16 mostram algumas destas funções.

Figura 11 – Página de gerenciamento de usuários

USUÁRIO	ENDEREÇO DE EMAIL	PRIMEIRO NOME	ÚLTIMO NOME	MEMBRO DA EQUIPE
adminlise	admin@lise.com	Admin		✓
nay	sfnayara@gmail.com	Nayara	Silva	✗
nayara.silva.f@hotmail.com	nayara.silva.f@hotmail.com	teste		✗

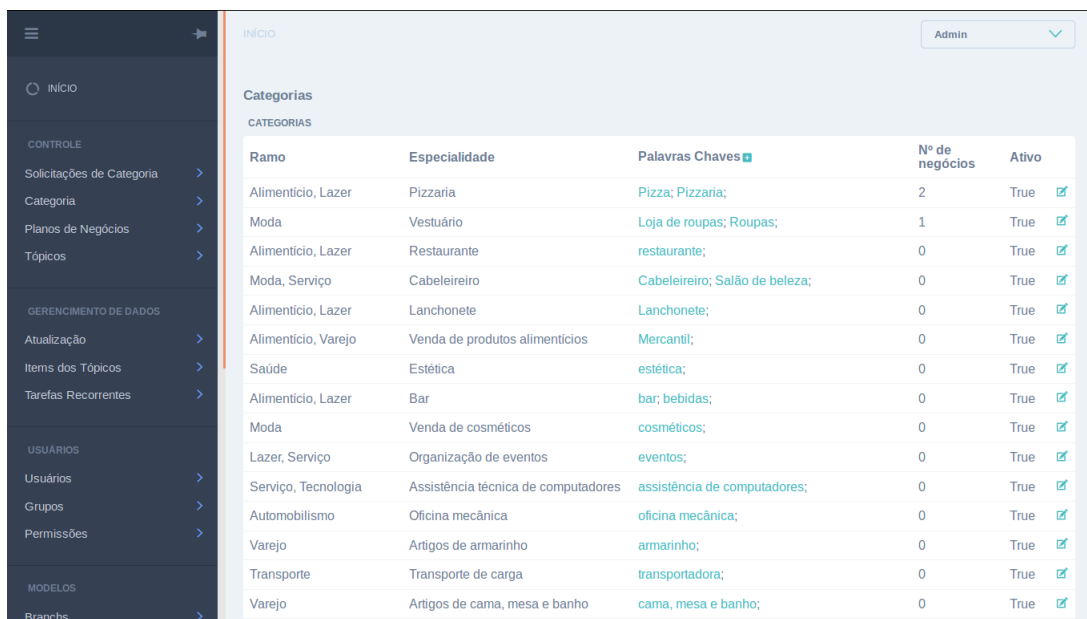
Fonte: Elaborada pela autora

Figura 12 – Página de gerenciamento de negócios planejados

Categoria	Cidade	Estado	Empreendedor	Dias de análise	Status
Vestuário	Santos	SP	Nayara	96	Ativo
Clínica de diagnóstico e consulta	Teresina	PI	Nayara	156	Ativo
Pizzaria	Manaus	AM	Nayara	96	Ativo
Pizzaria	Fortaleza	CE	teste	61	Solicitado

Fonte: Elaborada pela autora

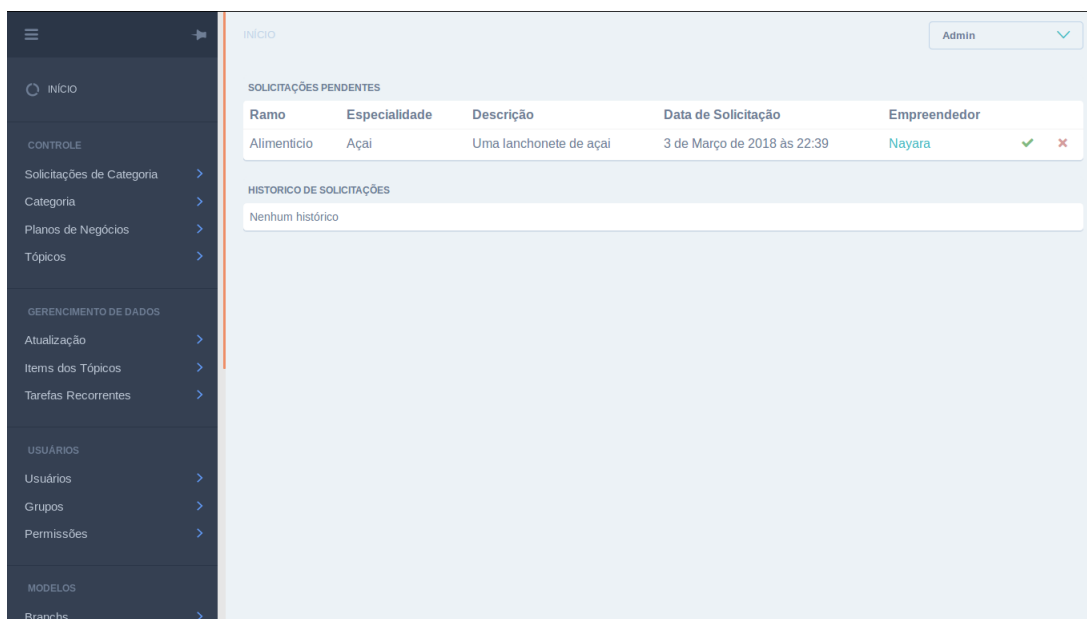
Figura 13 – Página de gerenciamento de categorias



Ramo	Especialidade	Palavras Chaves	Nº de negócios	Ativo
Alimentício, Lazer	Pizzaria	Pizza; Pizzaria;	2	True
Moda	Vestuário	Loja de roupas; Roupas;	1	True
Alimentício, Lazer	Restaurante	restaurante;	0	True
Moda, Serviço	Cabeleireiro	Cabeleireiro; Salão de beleza;	0	True
Alimentício, Lazer	Lanchonete	Lanchonete;	0	True
Alimentício, Varejo	Venda de produtos alimentícios	Mercantil;	0	True
Saúde	Estética	estética;	0	True
Alimentício, Lazer	Bar	bar; bebidas;	0	True
Moda	Venda de cosméticos	cosméticos;	0	True
Lazer, Serviço	Organização de eventos	eventos;	0	True
Serviço, Tecnologia	Assistência técnica de computadores	assistência de computadores;	0	True
Automobilismo	Oficina mecânica	oficina mecânica;	0	True
Varejo	Artigos de armarinho	armarinho;	0	True
Transporte	Transporte de carga	transportadora;	0	True
Varejo	Artigos de cama, mesa e banho	cama, mesa e banho;	0	True

Fonte: Elaborada pela autora

Figura 14 – Página de gerenciamento de solicitações de categorias



Ramo	Especialidade	Descrição	Data de Solicitação	Empreendedor
Alimentício	Açaí	Uma lanchonete de açaí	3 de Março de 2018 às 22:39	Nayara

HISTÓRICO DE SOLICITAÇÕES

Nenhum histórico

Fonte: Elaborada pela autora

Figura 15 – Página de gerenciamento de atualização de dados

INÍCIO

Admin

Gerenciamento de Dados

Atualizar Testar Conexão Conectar Cogroo

ATUALIZAÇÕES DE DADOS

Numero	Tempo inicial	Tempo final	Duração (s)	Planos	Status
106	6 de Março de 2018 às 11:16	None	---	0	Em andamento
105	6 de Março de 2018 às 04:07	None	---	0	Em andamento
104	6 de Março de 2018 às 03:03	6 de Março de 2018 às 03:16	761	3	Concluído
103	6 de Março de 2018 às 02:42	6 de Março de 2018 às 03:01	1155	3	Concluído
102	6 de Março de 2018 às 02:35	6 de Março de 2018 às 02:40	276	0	Falhado
101	6 de Março de 2018 às 02:28	6 de Março de 2018 às 02:33	288	0	Falhado
100	6 de Março de 2018 às 02:18	6 de Março de 2018 às 02:22	233	0	Falhado
99	6 de Março de 2018 às 02:16	None	---	0	Em andamento
98	6 de Março de 2018 às 01:57	None	---	0	Em andamento
97	6 de Março de 2018 às 01:53	6 de Março de 2018 às 01:56	164	0	Falhado
96	6 de Março de 2018 às 01:48	6 de Março de 2018 às 01:51	170	0	Falhado
95	6 de Março de 2018 às 01:38	6 de Março de 2018 às 01:42	220	0	Falhado
94	6 de Março de 2018 às 01:26	6 de Março de 2018 às 01:32	400	0	Falhado
93	6 de Março de 2018 às 01:08	6 de Março de 2018 às 01:18	593	0	Falhado
92	6 de Março de 2018 às 00:02	None	---	0	Em andamento

Fonte: Elaborada pela autora

Figura 16 – Página de gerenciamento de tarefas

INÍCIO > DJANGO Q > SCHEDULED TASKS

Admin

Pesquisar

Por Schedule Type Por Next Run

+ Adicionar Scheduled task

PESQUISAR

ID	NAME	FUNC	SCHEDULE TYPE	REPEATS	NEXT RUN	LAST RUN	SUCCESS
2	coleta e mineração de dados	managedata.mining.run_mining.update_business_plan	Daily	-2	6 de Março de 2018 às 11:30	[sixteen-winter-sodium-seven]	✓

0 de 1 selecionados

1 Scheduled task

Fonte: Elaborada pela autora

RESUMO

Neste capítulo, foram mostrados os resultados da verificação, onde testa a polaridade e a sumarização do Lise, comparando com as respostas de um humano. Estes resultados foram utilizados em testes de hipóteses, que confirmaram a hipótese nula apresentada no capítulo de

metodologia. Também foram apresentadas as principais telas do Lise que o empreendedor irá interagir e os *cards* que mostram as informações da análise de concorrentes.

6 CONCLUSÃO

Este trabalho apresentou um mecanismo para realizar a análise de concorrentes automatizada, onde o empreendedor conseguirá realizar esta etapa do plano de negócio com mais rapidez e menos trabalho que a forma manual. Para isso foi desenvolvida uma ferramenta online chamada Lise. A fonte de dados do Lise para coletar dados dos concorrentes foi o *Google Maps*, onde é possível saber as opiniões dos clientes dos estabelecimentos sem causar incômodo.

Para verificar se as análises de concorrentes realizadas no Lise são mesmo verídicas e consideráveis com o resultado manual, foi realizada uma verificação do sistema. Esta verificação utilizou uma amostra de avaliações sobre os concorrentes, onde o Lise e um humano deveriam analisar a polaridade e a sumarização das avaliações. Depois disto, as respostas foram comparadas, gerando os resultados mostrados no capítulo de resultados. Esta verificação utilizou testes de hipóteses para verificar a confiabilidade dos resultados. Com estes testes, a hipótese nula, que afirma que o Lise tem uma proporção de acertos maior ou igual a 75% sobre as análises das avaliações dos concorrentes, foi aceita tanto para a polaridade quanto para sumarização.

Para o empreendedor ter acesso a todas as informações sobre os concorrentes foi desenvolvido um sistema *web*. O Lise apresenta estas informações de forma atraente e fácil de entender, através de gráficos, tabelas, lista, entre outros recursos.

6.1 TRABALHOS FUTUROS

Mesmo considerando um bom resultado na verificação, o primeiro trabalho futuro seria melhorar o algoritmo de análise de texto para gerar resultados mais equivalentes com os de um humano. Isto poderia ser feito, realizando mais testes, analisando os resultados incorretos da primeira versão e até integrando ao sistema alguns conceitos de *Machine Learning*.

Outra mudança que poderia ser feita seria adicionar outra fonte de dados, além do *Google Maps*. Uma opção seria o *site* Apontador¹ que tem vários comentários relevantes e possui API para a coleta de dados. Mas, existem outros diversos *sites* de avaliações de empresas que poderiam ser úteis para análise de concorrentes.

Pensando nas empresas digitais, as quais o Lise ainda não atende, seria um avanço para o sistema conseguir realizar a análise de concorrentes para empreendedores interessados nas empresas digitais. Isto porque o número de empresas deste tipo está aumentando a cada ano. Mas isso seria outro desafio, pois as informações e a coleta seriam trabalhadas de outra forma. Uma sugestão de fonte de dados para isso seria o *site* Reclame Aqui², que contém muitas reclamações e também soluções de diversas empresas.

¹ <https://www.apontador.com.br/>

² <https://www.reclameaqui.com.br>

Por fim, outro desafio para trabalhos futuros seria automatizar o outro campo da análise de mercado: a análise de clientes. Também ele é consideravelmente importante para criar estratégias empresariais. Isto seria viável através de redes sociais. Como hoje muitas empresas participam das redes sociais, seria possível investigar os seus clientes na página social dos próprios clientes.

REFERÊNCIAS

- ARAÚJO, M. et al. Métodos para análise de sentimentos no twitter. In: *Proceedings of the 19th Brazilian symposium on Multimedia and the Web (WebMedia'13)*. [S.l.: s.n.], 2013. Citado na página 19.
- ARENAS, A. G. et al. Exploring machine translation on the web. *Tradumàtica: traducció i tecnologies de la informació i la comunicació*, n. 8, p. 1–6, 2010. Citado na página 51.
- BAEZA-YATES, R.; RIBEIRO-NETO, B. et al. *Modern information retrieval*. [S.l.]: ACM press New York, 1999. v. 463. Citado na página 17.
- BALAHUR, A.; TURCHI, M. Multilingual sentiment analysis using machine translation? In: ASSOCIATION FOR COMPUTATIONAL LINGUISTICS. *Proceedings of the 3rd workshop in computational approaches to subjectivity and sentiment analysis*. [S.l.], 2012. p. 52–60. Citado 2 vezes nas páginas 19 e 21.
- CHIAVENATO, I. *Empreendedorismo: dando asas ao espírito empreendedor*. [S.l.]: editora manole, 2004. Citado na página 14.
- DORNELAS, J. *COMO MONTAR UM PLANO DE NEGÓCIOS SIMPLES E PRÁTICO*. 2013. Disponível em: <<http://revistapegn.globo.com/Como-comecar/noticia/2013/06/como-montar-um-plano-de-negocio-simples-e-pratico.html>>. Acesso em: 10 fev. 2018. Citado na página 14.
- DORNELAS, J. C. A. Plano de negócios: seu guia definitivo. *Rio de Janeiro: Campus*, 2011. Citado na página 17.
- EGESTOR. *Por que devo fazer análise da concorrência?* 2017. Disponível em: <<https://blog.egestor.com.br/por-que-devo-fazer-analise-da-concorrencia>>. Acesso em: 10 fev. 2018. Citado na página 14.
- FILHO, N. C. *Elaboração de projetos empresariais: análise estratégica, estudo de viabilidade e plano de negócio*. [S.l.]: Atlas, 2009. Citado 2 vezes nas páginas 17 e 23.
- FRAKES, W. B.; BAEZA-YATES, R. *Information retrieval: Data structures & algorithms*. [S.l.]: prentice Hall Englewood Cliffs, New Jersey, 1992. v. 331. Citado na página 17.
- FREUND, J. E.; SIMON, G. A. Estatística aplicada. *Economia, Administração e*, 2009. Citado 4 vezes nas páginas 24, 25, 36 e 37.
- GIBICOSKI, S. D. *Etapas do plano de negócio para seu empreendimento*. 2017. Disponível em: <<http://www.administradores.com.br/artigos/carreira/etapas-do-plano-de-negocio-do-empreendimento/79440/>>. Acesso em: 13 jul. 2017. Citado na página 14.
- GOMES, I. M. Manual como elaborar um plano de marketing. *Belo Horizonte: Sebrae/MG*, v. 20005, 2005. Citado na página 20.
- HAN, J.; PEI, J.; KAMBER, M. *Data mining: concepts and techniques*. [S.l.]: Elsevier, 2011. Citado na página 18.

HUTTO, C. J.; GILBERT, E. Vader: A parsimonious rule-based model for sentiment analysis of social media text. In: *Eighth international AAAI conference on weblogs and social media*. [S.l.: s.n.], 2014. Citado 2 vezes nas páginas 19 e 22.

KINOSHITA, J.; SALVADOR, L.; MENEZES, C. Cogroo: a brazilian-portuguese grammar checker based on the cetenfolha corpus. In: *Proceedings of the 5th International Conference on Language Resources and Evaluation (LREC'2006), Genoa, Italy*. [S.l.: s.n.], 2006. Citado na página 22.

KURTZ, J. *Google Maps cria sistema de recompensa para aumentar o número de reviews*. 2017. Disponível em: <<http://www.techtudo.com.br/noticias/noticia/2015/02/google-maps-cria-sistema-de-recompensa-para-aumentar-o-numero-de-reviews.html>>. Acesso em: 19 ago. 2017. Citado na página 15.

LAUBE, K. P. *Entendendo o Django*. 2017. Disponível em: <<https://www.profissionaisti.com.br/2009/04/entendendo-o-django/>>. Acesso em: 11 ago. 2017. Citado na página 33.

LINOFF, G. S.; BERRY, M. J. *Data mining techniques: for marketing, sales, and customer relationship management*. [S.l.]: John Wiley & Sons, 2011. Citado na página 18.

LIU, B.; ZHANG, L. A survey of opinion mining and sentiment analysis. In: *Mining text data*. [S.l.]: Springer, 2012. p. 415–463. Citado na página 18.

LOPEZ, A. Statistical machine translation. *ACM Computing Surveys (CSUR)*, ACM, v. 40, n. 3, p. 8, 2008. Citado na página 51.

MELO, L. *Crise faz empreendedorismo por necessidade voltar a crescer no Brasil*. 2017. Disponível em: <<https://g1.globo.com/economia/pme/noticia/crise-faz-empreendedorismo-por-necessidade-voltar-a-crescer-no-brasil.ghtml>>. Acesso em: 19 fev. 2018. Citado na página 14.

MICRO, S. B. D. A. À. *Sobrevivência das empresas no brasil*. Brasília, DF: SEBRAE, 2013. Citado na página 15.

MIOT, H. A. Tamanho da amostra em estudos clínicos e experimentais. *J Vasc bras*, SciELO Brasil, v. 10, n. 4, p. 275–8, 2011. Citado 2 vezes nas páginas 23 e 24.

MOHAMMADIAN, M. *Intelligent agents for data mining and information retrieval*. [S.l.]: IGI Global, 2004. Citado na página 18.

OCHOA, C. *Qual é o tamanho da amostra que eu preciso?* 2013. Disponível em: <<https://www.netquest.com/blog/br/blog/br/qual-e-o-tamanho-de-amostra-que-preciso>>. Acesso em: 20 fev. 2018. Citado na página 23.

PANG, B.; LEE, L. et al. Opinion mining and sentiment analysis. *Foundations and Trends® in Information Retrieval*, Now Publishers, Inc., v. 2, n. 1–2, p. 1–135, 2008. Citado na página 19.

Q, D. *Brokers*. 2018. Disponível em: <<http://django-q.readthedocs.io/en/latest/brokers.html>>. Acesso em: 1 jan. 2018. Citado na página 22.

Q, D. *Welcome to Django Q*. 2018. Disponível em: <<http://django-q.readthedocs.io/en/latest/>>. Acesso em: 1 jan. 2018. Citado na página 22.

- REIS, E. et al. Estatística aplicada (vol. 2). *Lisboa: Edições Sílabo*, 1999. Citado 2 vezes nas páginas 23 e 24.
- REIS, J. C. et al. Uma abordagem multilíngue para análise de sentimentos. *IV Brazilian*, 2015. Citado 2 vezes nas páginas 19 e 22.
- ROSA, C. A. Como elaborar um plano de negócio. *SEBRAE/MG*, 2004. Citado na página 14.
- SOMMERVILLE, I. Engenharia de software-9ª edição. *Ed Person Education*, 2011. Citado na página 31.
- TOLEDO, M. *Plano de Negócios para Startups*. 2015. Citado na página 17.
- WAZLAWICK, R. *Análise e Projeto de Sistemas de Informação Orientados*. [S.l.]: Elsevier Brasil, 2010. v. 2. Citado na página 28.
- WU, Y. et al. Google's neural machine translation system: Bridging the gap between human and machine translation. *arXiv preprint arXiv:1609.08144*, 2016. Citado na página 51.
- XU, K. et al. Mining comparative opinions from customer reviews for competitive intelligence. *Decision support systems*, Elsevier, v. 50, n. 4, p. 743–754, 2011. Citado na página 23.

APÊNDICE A – SELEÇÃO DE TRADUTOR PARA O LISE

Considerando a qualidade da tradução e a avaliação feita por Arenas (2010), foram analisados alguns pontos dos serviços de tradução da *Microsoft* e *Google*. Ambos usam a abordagem *Neural Machine Translation* (NMT) que realiza traduções através de uma rede neural, baseada no *Statistical Machine Translation* (SMT). Os algoritmos SMT estão avançando cada vez mais, gerando diversas utilidades. Eles aprendem automaticamente a traduzir, depois de examinar diversas traduções feitas por humanos (LOPEZ, 2008). O *Google* aprimorou ainda mais a técnica de NMT criando o *Google Neural Machine Translation* (GNMT), utilizada no serviço dele (WU et al., 2016).

A Tabela 6 mostra os pontos analisados nos dois serviços para utilização no Lise. No preço está o valor cobrado pelo serviço. O meio de uso se trata de como a aplicação do Lise iria utilizar o serviço. O método de tradução diz de que maneira o serviço traduz os textos. Em plataforma *Python* é respondido se tem biblioteca dos tradutores na linguagem *Python*. O ponto de ação após palavra desconhecida informa o que o serviço faz ao não reconhecer a palavra. Também é analisado se os serviços traduzem do português ao inglês, ou vice-versa. Por último é mostrada a tecnologia utilizada.

Tabela 6 – Pontos analisados dos tradutores

Tradutor	Microsoft Translator	Google Translation
Preço	\$10 (Por milhão de caracteres) Dois milhões de caracteres grátis	\$20 (Por milhão de caracteres)
Meio de uso	API REST	API REST
Método de tradução	Através de marcação HTML	Tradução direta de texto
Plataforma Python	Sim	Sim
Ação após palavra desconhecida	Remove palavra	Permanece na linguagem original
Traduzir em português e inglês	Sim	Sim
Tecnologia	NMT	GNMT

Fonte: Elaborada pela autora

Considerando as informações da Tabela 6, o serviço de tradução escolhido foi do *Google*. Mesmo sendo mais caro, ele fornece mais qualidade e segurança no resultado da tradução. Como a etapa de tradução influencia bastante no resultado final das análises, a qualidade deste serviço obteve prioridade.

APÊNDICE B – ESPECIFICAÇÃO DE CASOS DE USO

CASO DE USO 1

- **Caso de uso:** Analisar concorrentes .
- **Atores:** Empreendedor.
- **Precondições:** O empreendedor deve estar logado no sistema e ter pelo menos um negócio cadastrado.
- **Pós-condições:** Disponibilidade dos resultados da análise de concorrentes do determinado negócio do empreendedor.
- **Requisitos correlacionados:** R1 e R9.
- **Fluxo principal:**
 1. O empreendedor entra no sistema;
 2. O empreendedor escolhe um dos seus negócios cadastrados para visualizar;
 3. O empreendedor abre o painel de controle do negócio escolhido;
 4. O empreendedor ver os resultados da análise de concorrentes.
- Tratamento de exceções:
 - 3a. O negócio está aguardando a resposta da solicitação de categoria.**
 - 3a.1** O empreendedor sai do sistema;
 - 3a.2** O empreendedor volta depois;
 - 3a.3** Retorna ao fluxo principal no passo 1.
 - 4a. O negócio ainda não teve sua primeira análise.**
 - 4a.1** O empreendedor sai do sistema;
 - 4a.2** O empreendedor volta depois;
 - 4a.3** Retorna ao fluxo principal no passo 1.

CASO DE USO 2

- **Caso de uso:** Coletar dados dos concorrentes.
- **Atores:** Google Maps.
- **Interessados:** Empreendedor.
- **Precondições:** O negócio e sua categoria deverão estar cadastrados.

- **Pós-condições:** Os dados coletados serão guardados na base de dados.
- **Requisitos correlacionados:** R6.
- **Fluxo principal:**
 1. O sistema procura negócio cadastrado;
 2. O sistema verifica as especificações do negócio, que foi incluído na sua categoria;
 3. O sistema se conecta com o Google Maps;
 4. O sistema pesquisa os concorrentes no Google Maps;
 5. O sistema coleta os dados encontrados no Google Maps;
 6. Os dados são adicionados na base de dados.

CASO DE USO 3

- **Caso de uso:** Cadastrar negócio planejado.
- **Atores:** Empreendedor e Administrador.
- **Precondições:** O empreendedor deverá estar logado no sistema e ter alguma ideia de negócio de loja física.
- **Pós-condições:** Atualização nos registros de planos de negócios do empreendedor.
- **Requisitos correlacionados:** R2, R3 , R7, R9 e R10.
- **Fluxo principal:**
 1. O empreendedor realiza login no sistema;
 2. O empreendedor clica na opção de Cadastrar plano de negócio;
 3. O empreendedor preenche os campos do cadastro: localidade, categoria e descrição.
 4. O empreendedor envia o cadastro;
 5. O negócio é criado;
 6. O empreendedor é notificado.
- **Tratamento de exceções:**
 - 4a. Não há categoria desejada pelo empreendedor.**
 - 4a.1** O empreendedor aperta a opção de solicitar categoria no campo de tipo de categoria;
 - 4a.2** O empreendedor descreve a categoria desejada;

4a.3 Retorna fluxo principal no passo 4.

6a. Há solicitação de categoria.

6a.1 O administrador recebe o pedido de categoria;

6a.2 O administrador avalia o pedido de categoria;

6a.3 O administrador cadastra a categoria;

6a.4 Retorna ao fluxo principal no passo 6.

6b. Categoria solicitada é inviável para o sistema.

6b.1 O administrador recusa a solicitação da categoria;

6b.2 O negócio não é criado;

6b.3 O empreendedor é notificado sobre o problema.

6c. Categoria solicitada está mal descrita.

6c.1 O administrador solicita uma nova descrição da categoria ao empreendedor;

6c.2 O negócio não é criado;

6c.3 O empreendedor é notificado sobre o problema.

CASO DE USO 4

- **Caso de uso:** Gerenciar categorias de negócios.
- **Atores:** Administrador.
- **Interessados:** Empreendedor.
- **Precondições:** O administrador deverá ter alguma mudança de categoria para realizar.
- **Pós-condições:** Atualização na lista de categorias de negócios do sistema.
- **Requisitos correlacionados:** R3 e R7.
- **Fluxo principal:**
 1. O administrador inicia a mudança devida na lista de categorias: adicionar, excluir, atualizar ou visualizar;
 2. O administrador termina a mudança;
 3. A lista de categorias de negócios é atualizada.

CASO DE USO 5

- **Caso de uso:** Relacionar palavras das opiniões.

- **Atores:** Administrador.
- **Precondições:** O sistema identificou novas palavras durante a atualização de dados.
- **Pós-condições:** Modificação nas informações com dados novos.
- **Requisitos correlacionados:** R8.
- **Fluxo principal:**
 1. O administrador visualiza as novas palavras;
 2. O administrador analisa cada uma;
 3. O administrador adiciona um tópico e as categorias em cada palavra;
- **Tratamento de exceções:**
 - 3a. A palavra não tem relação a nenhum tópico.**
 - 3a.1** O empreendedor aperta a opção de recusar palavra;
 - 3a.2** O empreendedor seleciona outra palavra;
 - 3a.3** Retorna fluxo principal no passo 3.
 - 3b. A palavra está relacionada, mas ele não está no sistema.**
 - 3b.1** O empreendedor aperta a opção de criar um novo tópico;
 - 3b.2** O empreendedor descreve o nome e a descrição no formulário de criação de novo tópico;
 - 3b.3** O empreendedor salva o novo tópico;
 - 3b.4** O empreendedor relaciona a palavra com o novo tópico;
 - 3b.5** O empreendedor seleciona outra palavra;
 - 3b.6** Retorna fluxo principal no passo 3.

APÊNDICE C – LISTA DE CATEGORIAS INICIAL DO LISE

Ramos	Categoria	Palavras Chaves
Alimentício Lazer	Pizzaria	Pizza Pizzaria
Moda	Vestuário	Loja de roupas Roupas
Alimentício Lazer	Restaurante	restaurante
Moda Serviço	Cabeleireiro	Cabeleireiro Salão de beleza
Alimentício Lazer	Lanchonete	Lanchonete
Alimentício Varejo	Venda de produtos alimentícios	Mercantil
Saúde	Estética	estética
Alimentício Lazer	Bar	bar bebidas
Moda	Venda de cosméticos	cosméticos
Lazer Serviço	Organização de eventos	eventos
Serviço Tecnologia	Assistência técnica de computadores	assistência de computadores
Automobilismo	Oficina mecânica	oficina mecânica
Varejo	Artigos de armarinho	armarinho
Transporte	Transporte de carga	transportadora
Varejo	Artigos de cama, mesa e banho	cama, mesa e banho
Tecnologia Varejo	Venda de equipamentos de informática	venda informática
Ensino	Atividade de ensino	ensino
Moda	Venda de acessórios de beleza	moda e acessórios
Transporte	Taxi	taxi
Arte e cultura	Produção fotográfica	fotografia
Automobilismo Serviço	Lanternagem, funilaria e pintura de automóveis	funilaria de veículos lanternagem de veículos pintura de veículos;
Automobilismo	Peças e acessórios de veículos	peças e acessórios de veículos
Construção	Venda de materiais de construção	casa e construção materiais de construção
Empresarial Serviço	Marketing direto	marketing
Moda	Venda de calçados	loja de calçados
Varejo	Venda de variedades	loja de variedades
Empresarial Serviço	Treinamento de desenvolvimento pessoal e gerencial	recursos humanos rh

Ramos	Categoria	Palavras Chaves
Cuidados Serviço	Loja de animais	loja de animais pet shop
Empresarial, Serviço	Atividades de contabilidade	contabil contabilidade
Arte e cultura Varejo	Equipamento musical	equipamentos musical instrumento musical
Alimentício	Padaria e confeitaria	confeitaria padaria panificadora
Arte e cultura Ensino	Ensino de arte e cultura	arte marciais ensino arte e cultura ensino de dança escola de teatro
Alimentício	Açougue	figrorífico
Assistência Serviço	Assistência em equipamentos eletroeletrônicos	assistencia de eletronicos
Moda Varejo	Artigos de joalheria e bijuterias	bijuteria joalheria
Lazer Serviço	Agência de viagens	agencia de turismo agencia de viagens
Saúde	Produtos farmacêuticos	drogaria farmacia
Saúde	Clinica de diagnóstico e consulta	clinica de diagnostico consulta médica diagnostico
Construção	Loja de ferragens e ferramentas	loja de ferragens loja de ferramentas
Varejo	Artigos de papelaria	artigos de papelaria papelaria
Assistência Empresarial Serviço	Consultoria em gestão empresarial	assessoria de empresas consultoria empresariais gestão empresarial
Varejo	Lojas de departamentos	loja de departamento loja de magazine
Varejo	Loja de eletrodomésticos	loja de eletrodomésticos
Serviço Transporte	Serviço de entrega rapida	entrega rápida serviço de entrega rapida
Serviço	Serviços de engenharia	engenharia serviço de engenharia
Moda Saúde	Artigos de óptica	óptica ótica
Serviço	Empreendimentos imobiliários	empreendimentos imobiliários imobiliaria
Alimentício	Lojas de doces e bombons	doceria loja de balas loja de bombons loja de doces

Ramos	Categoria	Palavras Chaves
Saúde Serviço	Serviço de condicionamento físico	academia condicionamento fisico ginastica pilates
Serviço	Serviços advocatícios	advocacia advogado serviço de advogados
Lazer Serviço	Hoteis e pousadas	hospedagem hoteis pousada
Varejo	Loja de brinquedos	brinquedos loja de brinquedos
Construção Varejo	Loja de vidros	loja de vidros vidraçaria vidraceiro
Varejo	Loja de plantas e flores	floricultura loja de flores loja de plantas
Saúde Serviço	Serviço odontológico	dentista odontologia
Varejo	Artigos esportivos	artigos esportivos equipamentos esportivos materiais esportivos
Serviço	Lavanderia	lavanderia serviço de lavanderia
Construção	Madeira e artefatos	madeira madeira e artefatos
Arte e cultura Ensino Serviço	Ensino de idiomas	ensino de idiomas escola de idiomas
Varejo	Vendas de livros	livraria venda de livros

ANEXO A – TABELA DE DISTRIBUIÇÃO NORMAL PADRÃO

DISTRIBUIÇÃO NORMAL PADRÃO

VALORES DA FUNÇÃO DE DISTRIBUIÇÃO

$$F(z) = \int_{-\infty}^z \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt$$

z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5753
0,2	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141
0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480	0,6517
0,4	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879
0,5	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,7224
0,6	0,7257	0,7291	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7517	0,7549
0,7	0,7580	0,7611	0,7642	0,7673	0,7704	0,7734	0,7764	0,7794	0,7823	0,7852
0,8	0,7881	0,7910	0,7939	0,7967	0,7995	0,8023	0,8051	0,8078	0,8106	0,8133
0,9	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315	0,8340	0,8365	0,8389
1,0	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554	0,8577	0,8599	0,8621
1,1	0,8643	0,8665	0,8686	0,8708	0,8729	0,8749	0,8770	0,8790	0,8810	0,8830
1,2	0,8849	0,8869	0,8888	0,8907	0,8925	0,8944	0,8962	0,8980	0,8997	0,9015
1,3	0,9032	0,9049	0,9066	0,9082	0,9099	0,9115	0,9131	0,9147	0,9162	0,9177
1,4	0,9192	0,9207	0,9222	0,9236	0,9251	0,9265	0,9279	0,9292	0,9306	0,9319
1,5	0,9332	0,9345	0,9357	0,9370	0,9382	0,9394	0,9406	0,9418	0,9429	0,9441
1,6	0,9452	0,9463	0,9474	0,9484	0,9495	0,9505	0,9515	0,9525	0,9535	0,9545
1,7	0,9554	0,9564	0,9573	0,9582	0,9591	0,9599	0,9608	0,9616	0,9625	0,9633
1,8	0,9641	0,9649	0,9656	0,9664	0,9671	0,9678	0,9686	0,9693	0,9699	0,9706
1,9	0,9713	0,9719	0,9726	0,9732	0,9738	0,9744	0,9750	0,9756	0,9761	0,9767
2,0	0,9772	0,9778	0,9783	0,9788	0,9793	0,9798	0,9803	0,9808	0,9812	0,9817
2,1	0,9821	0,9826	0,9830	0,9834	0,9838	0,9842	0,9846	0,9850	0,9854	0,9857
2,2	0,9861	0,9864	0,9868	0,9871	0,9875	0,9878	0,9881	0,9884	0,9887	0,9890
2,3	0,9893	0,9896	0,9898	0,9901	0,9904	0,9906	0,9909	0,9911	0,9913	0,9916
2,4	0,9918	0,9920	0,9922	0,9925	0,9927	0,9929	0,9931	0,9932	0,9934	0,9936
2,5	0,9938	0,9940	0,9941	0,9943	0,9945	0,9946	0,9948	0,9949	0,9951	0,9952
2,6	0,9953	0,9955	0,9956	0,9957	0,9959	0,9960	0,9961	0,9962	0,9963	0,9964
2,7	0,9965	0,9966	0,9967	0,9968	0,9969	0,9970	0,9971	0,9972	0,9973	0,9974
2,8	0,9974	0,9975	0,9976	0,9977	0,9977	0,9978	0,9979	0,9979	0,9980	0,9981
2,9	0,9981	0,9982	0,9982	0,9983	0,9984	0,9984	0,9985	0,9985	0,9986	0,9986
3,0	0,9987	0,9987	0,9987	0,9988	0,9988	0,9989	0,9989	0,9989	0,9990	0,9990
3,1	0,9990	0,9991	0,9991	0,9991	0,9992	0,9992	0,9992	0,9992	0,9993	0,9993
3,2	0,9993	0,9993	0,9994	0,9994	0,9994	0,9994	0,9994	0,9995	0,9995	0,9995
3,3	0,9995	0,9995	0,9995	0,9996	0,9996	0,9996	0,9996	0,9996	0,9996	0,9997
3,4	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9998

z	1,282	1,645	1,960	2,326	2,576	3,090	3,291	3,891	4,417
F(z)	0,90	0,95	0,975	0,99	0,995	0,999	0,9995	0,99995	0,999995
2[1 - F(z)]	0,20	0,10	0,05	0,02	0,01	0,002	0,001	0,0001	0,00001