



Adilaraima Martinez Barrio

**VISUALIZAÇÃO DE TEXTO
MINERADO PARA EXTRAÇÃO DE REQUISITOS**

Dissertação de Mestrado

Dissertação apresentada como requisito parcial para obtenção do grau de Mestre pelo Programa de Pós-graduação em Informática da PUC-Rio.

Orientador: Prof. Julio Cesar Sampaio do Prado Leite

Rio de Janeiro
Março de 2020



Adilaraima Martinez Barrio

VISUALIZAÇÃO DE TEXTO MINERADO PARA EXTRAÇÃO DE REQUISITOS

Dissertação apresentada como requisito parcial para obtenção do grau de Mestre pelo Programa de Pós-Graduação em Informática do Centro Técnico Científico da PUC-Rio. Aprovada pela Comissão Examinadora abaixo.

Prof. Julio Cesar Sampaio do Pardo Leite
Orientador
Departamento de Informática – PUC-Rio

Prof^a. Vera Maria Werneck Benjamin
Departamento de Informática – UERJ

Prof. Marco Antonio Casanova
Departamento de Informática – PUC-Rio

Todos os direitos reservados. É proibida a reprodução total ou parcial do trabalho sem autorização da universidade, da autora e do orientador.

Adilaraima Martinez Barrio

Atua no mercado de Engenharia de Software desde 2010. Graduiu-se em Engenharia Informática pela Universidade de Ciências Informáticas da Havana Cuba em 2010. Ingressou no Mestrado em Informática na Pontifícia Universidade Católica do Rio de Janeiro em 2018.

Ficha Catalográfica

Martinez Barrio, Adilaraima

Visualização de texto minerado para a extração de requisitos / Adilaraima Martinez Barrio; orientador: Julio Cesar Sampaio do Prado Leite. – 2020.

108 f. : il. color. ; 30 cm

Dissertação (mestrado)–Pontifícia Universidade Católica do Rio de Janeiro, Departamento de Informática, 2020.

Inclui bibliografia

1. Informática – Teses. 2. Visualização. 3. Mineração de texto. 4. Engenharia de requisitos. 5. Biblioteca digital. I. Leite, Julio Cesar Sampaio do Prado. II. Pontifícia Universidade Católica do Rio de Janeiro. Departamento de Informática. III. Título.

CDD: 004

A minhas sobrinhas, meu sol e a minha lua.

Agradecimentos

Aos meus pais, pela educação, atenção e carinho de todas as horas.

Ao meu orientador Professor Júlio Leite pelo estímulo, conhecimentos, dedicação e paciência para a realização deste trabalho. É uma pessoa incrível!!

Ao meu complemento, amigo, confidente de horas difíceis Maikel Navarro pela força e o amor que me dá dia a dia.

Aos meus familiares de Cuba e de Brasil que torceram muito por meu sucesso.

Aos meus amigos cubanos que já formam parte da minha família.

Ao meu amigo Jonathan Petzold por todo o apoio e as horas de revisões do meu português.

Aos meus colegas da PUC-Rio, do Tecgraf e o grupo de pesquisa que de uma forma ou de outra me estimularam a chegar até aqui.

A todos os professores e funcionários do Departamento pelos ensinamentos e pela ajuda, em especial ao que gentilmente aceitaram o convite para pertencer a minha banca.

A PUC-Rio, pelos auxílios concedidos, sem os quais este trabalho não poderia ter sido realizado.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001.

Obrigada a todos que me apoiaram esse tempo todo!!!

Resumo

Barrio, Adilaraima Martinez; Leite, Julio Cesar Sampaio do Prado (Orientador). **Visualização de texto minerado para extração de requisitos**. Rio de Janeiro, 2020. 108. Dissertação de Mestrado – Departamento de Informática, Pontifícia Universidade Católica do Rio de Janeiro.

Com os avanços de novas tecnologias de desenvolvimento de software, a mineração de texto tem ganho protagonismo na área de Engenharia de Requisitos (ER), já que a rede (Web) possibilita o acesso a grandes quantidades de informação. A utilização de técnicas de visualização ganha importância nesse sentido, porque permite agilizar a descoberta de conhecimento com visualizações adequadas. Neste trabalho disponibiliza-se uma estratégia para que o engenheiro de requisitos consiga acessar às visualizações desenvolvidas por pesquisas no uso de mineração para elicitação de requisitos. Uma revisão da literatura possibilitou a proposta de classificação de tarefas de mineração e visualizações associadas, especialmente na ER. Esta classificação constitui a base de conhecimento de um software (Biblioteca Digital) que organiza e filtra informações de acordo com a inter-relação entre tarefas e categorias, mostrando as pesquisas que sustentam cada relação. Para a avaliação da estratégia, com foco no software, foi executado uma avaliação que mostra o potencial da abordagem para agilizar a obtenção de conhecimentos por parte do engenheiro de requisitos.

Palavras-chave

Visualização; Mineração de texto; Engenharia de requisitos; Biblioteca digital.

Abstract

Barrio, Adilaraima Martinez; Leite, Julio Cesar Sampaio do Prado (Advisor). **Text mining visualization for requirements extraction**. Rio de Janeiro, 2020. 108. Dissertação de Mestrado – Departamento de Informática, Pontifícia Universidade Católica do Rio de Janeiro.

With advances in new technologies available for software development, text mining has grown in importance in the area of Requirements Engineering (ER) due to the availability of large amounts of information on the Web. The use of visualization techniques gains importance in this regard, since it allows the speed-up of knowledge discovery with appropriate visualizations. This work provides a strategy for the requirements engineer to gain access, in an organized manner, to results of research in text mining with visualization in ER. For implementation, the classification of mining tasks and the visualizations categories in ER were summarized from the literature. This classification forms the knowledge base of a software (Digital Library) that organizes and filters information according to the interrelation between tasks and categories, showing the research that supports each relationship. For the evaluation of the strategy, with a focus on software, an evaluation was carried out that shows the potential of the approach to expedite the knowledge engineer's requirements.

Keywords

Visualization; Text mining; Requirements engineering; Digital library.

Sumário

1	INTRODUÇÃO	12
1.1	Motivação	12
1.2	Background	13
1.3	Visualização	14
1.4	Definição do problema	16
1.5	Organização do documento	17
1.6	Sumário	17
2	REVISÃO DE LITERATURA	18
2.1	Visualização da mineração de texto para ER	18
2.2	Tarefas analíticas	22
2.3	Categorias de visualização	24
2.4	Sumário	27
3	CORRELAÇÃO DAS TAREFAS DE MINERAÇÃO DE TEXTO COM CATEGORIAS DE VISUALIZAÇÃO PARA ER	28
3.1	Componentes	28
3.1.1.	Dados das tarefas analíticas	32
3.1.2.	Relação entre tarefa analítica e categoria de visualização	33
3.2	Sumário	37
4	BIBLIOTECA DIGITAL DE APOIO	40
4.1	Definição	40
4.2	Funcionalidade	41
4.3	Arquitetura da <i>MinViewREDL</i>	43
4.4	Uso da <i>MinViewREDL</i>	46
4.5	Sumário	48

5	AVALIAÇÃO DA ESTRATÉGIA	49
5.1	Objetivo	49
5.2	Seleção dos participantes	51
5.3	Execução	52
5.4	Análise dos resultados	56
5.5	Sumário	62
6	CONCLUSÃO	64
6.1	Resumo	64
6.2	Contribuições	65
6.3	Trabalhos futuros	66
7	REFERÊNCIAS BIBLIOGRÁFICAS	67
	APÊNDICE	70
	Apêndice I. Artigos da primeira etapa de revisão de literatura	70
	Apêndice II. Artigos de visualização na mineração de texto na ER	99
	Apêndice III. Conteúdos de ajuda para avaliação	103
	Apêndice IV. Avaliação aplicada	104
	Apêndice V. Estórias de Usuário	106

Lista de Figuras

Figura 1 Categoria de Visualização: <i>Distribution</i> .	25
Figura 2 Categoria de Visualização: <i>Correlation</i> .	25
Figura 3 Categoria de Visualização: <i>Ranking</i> .	25
Figura 4 Categoria de Visualização: <i>Part of Whole</i> .	25
Figura 5 Categoria de Visualização: <i>Evolution</i> .	26
Figura 6 Categoria de Visualização: <i>Map</i> .	26
Figura 7 Categoria de Visualização: <i>Flow</i> .	26
Figura 8 Estratégia para Integrar Mineração de texto com Visualização.	29
Figura 9 Modelo da inter-relação dos componentes.	30
Figura 10 Realçado em documento. Ref. Apêndice II No. [9].	33
Figura 11 Resumo de documento. Ref. Apêndice I No. [61].	34
Figura 12 Agrupamento de frases. Ref. Apêndice II No. [2].	34
Figura 13 Similaridade em frases. Ref. Apêndice II No. [6].	35
Figura 14 Rotulado em sentenças. Ref. Apêndice II No. [8].	36
Figura 15 Rotulado de documento. Ref. Apêndice II No. [14].	37
Figura 16 Arquitetura de Django (DJANGO SOFTWARE FOUNDATION, 2019).	44
Figura 17 Diagrama de componentes da Biblioteca Digital.	45
Figura 18 Modelo Entidade-Relação.	45
Figura 19 Interface para adicionar artigos novos.	46
Figura 20 Interface inicial da Biblioteca Digital.	47
Figura 21 Buscar artigo.	48
Figura 22 Detalhes do artigo.	48
Figura 23 Caracterização pelos anos de experiência.	52
Figura 24 Processo de avaliação.	53
Figura 25 Subprocesso de Avaliação Piloto.	53
Figura 26 Subprocesso do Avaliação Final.	55
Figura 27 Estatísticas para a pergunta 3.	58
Figura 28 Estatísticas para a pergunta 4.	58
Figura 29 Estatísticas para a pergunta 5.	58
Figura 30 Estatísticas para a pergunta 1.	59
Figura 31 Estatísticas para a pergunta 2.	60
Figura 32 Estatísticas para a pergunta 6.	61

Lista de Tabelas

Tabela 1 Tipo de dados por categorias.	31
Tabela 2 Associação entre Tarefa e Categoria.	38
Tabela 3 Questionário de avaliação.	50
Tabela 4 Resumo de resultados dos aspetos sobre o conhecimento adequado.	59
Tabela 5 Resumo de resultados da Operabilidade.	60

1 Introdução

1.1 Motivação

Software tornou-se um elemento fundamental da economia moderna em muitos países. Cada vez mais, novos componentes são adotados entre os produtos ou serviços usados diariamente (KHAN et al., 2015; LEITE, 2007). O processo de desenvolvimento desses produtos pode ser bem-sucedido com um bom entendimento, nas etapas iniciais, do contexto e do domínio onde o software vai ser usado e das necessidades próprias dos usuários que farão uso dele. Corresponde à Engenharia de Requisitos (ER), o processo de aquisição desse conhecimento.

Para a ER, a aplicação de várias técnicas para interpretar a linguagem humana (PORTUGAL, 2016) é primordial para conseguir o entendimento das necessidades do usuário. Eventualmente, técnicas foram desenvolvidas para vincular a mineração de texto na área, considerando que muita informação é proporcionada na elicitación, em formato textual (TAN, 1999).

Em projetos de software o fator tempo se torna uma variável importante devido ao aumento da competitividade nesse mercado. O engenheiro de requisitos precisa utilizar recursos que agilizem suas tarefas, sem afetar a qualidade do produto gerado. Assim técnicas de mineração de texto têm sido usadas para acelerar o processamento das informações na elicitación e principalmente para agilizar a leitura dos documentos (LEITE, 2007).

A principal diferença entre o processo de mineração de dados de forma tradicional e a mineração de textos é que, enquanto a abordagem convencional trabalha exclusivamente com dados estruturados, a mineração de textos lida com dados em linguagem natural e que, portanto, possui pouca ou nenhuma estrutura (SANTOS et al., 2015). Dessa forma acrescenta valor ao tratar com informações irrelevantes e ruidosas que devem ser removidas (SHIVAPRASAD, REDDY, 2016).

Após realizada uma mineração (tanto de dados quanto de texto), a exploração dos resultados desde diferentes perspectivas, é fundamental para ajudar aos usuários a entender as informações implícitas que podem existir. Dessa forma, é importante realçar que, se a mineração (de forma geral) ajuda a agilizar tarefas com grandes quantidades de dados, a visualização certa permitirá ganhar em entendimento (DIEHL, 2002).

Diante disso, destacamos que a realização da mineração de texto utilizando técnicas de visualização pode auxiliar na detecção de relacionamentos, especialmente nas atividades de elicitación de requisitos, especificamente na coleta de informações (fatos) a partir da automatização na leitura dos documentos (LEITE, 2007). As técnicas de visualização são mais do que um método de computação, já que contribuem para a transformação de informações em formato visual.

Mostrar de uma forma apropriada os resultados obtidos ainda é uma lacuna de pesquisa para mineração de texto na ER, principalmente nas atividades de elicitación (ABAD, NOAEEN, RUHE, 2016; FEATHER et al., 2006; PORTUGAL, 2016). Apesar da grande quantidade de pesquisas abordando técnicas de visualizações para texto minerado (KUCHER, KERREN, 2015) a engenharia de requisitos tem um baixo número de trabalhos tratando a forma de visualizar as informações textuais obtidas (GOTEL, MARCHESE, MORRIS, 2007).

Dessa forma, é um desafio obter meios que ajudem a engenheiros de requisitos, pesquisadores ou profissionais escolherem a visualização adequada para entender as informações obtidas através da mineração textual.

1.2 Background

A compreensão dos objetivos dos interessados do software deve ser promovida desde os estágios iniciais da sua construção. Esse entendimento tende a diminuir o tempo das iterações e até mesmo a quantidade de mudanças solicitadas pelos usuários finais, promovendo também uma maior gestão dos custos de evolução do software (DIEHL, 2002). Motivo suficiente para que a disciplina de ER tenha sido mais do que influente para o sucesso de qualquer sistema.

A ER tem como pilar fundamental de suas atividades enfrentar desafios principalmente nas atividades de elicitaco, ao enfrentar o aumento da informao na Internet. Embora estudos indiquem que 80% das informao de uma empresa esto contidas em documentos de texto (TAN, 1999), essas fontes apresentam diferentes formatos: textos, grficos, imagens e outros. Portanto, discernir entre as informao relevantes ainda  complicado. Torna-se um imensurvel excesso de trabalho, recursos e tempo para a explorao das mesmas.

O apoio da minerao de texto para melhorar a obteno de conhecimento e acelerar as tarefas de elicitaco  considerada uma soluo vivel e em ascenso (KUCHER, KERREN, 2015). A minerao de texto pode ser definida como uma ferramenta que vem da rea de processamento automtico de texto e que permite localizar e extrair as informao mais importantes e essenciais dos documentos (FENG, CHIAM, LO, 2017), alm de informao e conhecimento implcito e oculto (TAN, 1999) em grande corpora textual, estruturados ou no estruturados, como mensagens de e-mail, discursos, artigos, entre outros (ZAYAS, MEDEROS, 2007). Com essa premissa a minerao influencia nas atividades de elicitaco pois contribui na aquisio de conhecimento automaticamente, conseqentemente na construo de sistemas teis e de qualidade.

No entanto, a minerao de texto por si so no fornece muito valor para entender um contexto especfico; pode ser representada em diferentes formatos. Esse resultado, geralmente informao sem estrutura, pode ser difcil de entender assim como  difcil enxergar os relacionamentos que podem existir entre os dados. A ajuda da visualizao permite entender e processar melhor a informao ou o texto mais rapidamente.

1.3 Visualizao

O uso da visualizao atrai ateno de pesquisadores ou profissionais em geral que procuram apoiar pesquisas e agilizar as deciso de negcios com base em informao visuais enriquecidas (SHNEIDERMAN, 2012). O uso de visualizao  uma necessidade para acelerar o entendimento das informao (GOTEL, MARCHESE, MORRIS, 2007). Tambm  essencial resolver os

conflitos entre as partes interessadas e gerenciar as mudanças nos requisitos (ABAD, NOAEEN, RUHE, 2016).

Considera-se parte das atividades relacionadas à aprendizagem, o uso de representações apropriadas para extrair conhecimento (DUARTE et al., 2012). O objetivo da visualização, de maneira geral, é possibilitar que a informação seja apresentada através de formas simples e intuitivas, facilitando a assimilação da informação através do sentido da visão de um indivíduo (GOTEL, MARCHESE, MORRIS, 2007; WNUK et al., 2016).

Tradicionalmente a visualização na ER estava ligada à estrutura e aos relacionamentos (i), ao suporte para elicitación (ii) e em grupo de modelos (iii) (GOTEL, MARCHESE, MORRIS, 2007; WNUK et al., 2016). Argumentando em cada caso: (i) destacam-se visualizações nas representações hierárquicas dos documentos de requisitos, os relacionamentos de rastreabilidade entre os próprios requisitos e com outros artefatos. É realçada a aquisição de conhecimento em (ii) pelo uso de protótipos ou esboços que podem ser usados para explorar as necessidades das partes interessadas. Por outro lado, o uso dos modelos (iii), se manifesta de acordo com a finalidade desejada (intencional, processos, conceptual, orientado ao objeto ou serviço), inclusive ajudam na compreensão das atividades seguintes da ER. É importante destacar que a visualização também se manifesta no uso de ferramentas que auxiliam no gerenciamento de mudanças e, de maneira geral, no gerenciamento de requisitos.

As técnicas de visualização podem melhorar o processo de ER e, portanto, fornecer soluções práticas para reduzir mal-entendidos e lacunas de comunicação relacionadas à identificação e comunicação de requisitos (ABAD, NOAEEN, RUHE, 2016; FENG, CHIAM, LO, 2017; REDDIVARI, CHEN, NIU, 2012). Garantir o uso de visualizações na engenharia de requisitos é crucial (ABAD, NOAEEN, RUHE, 2016), e pode ser eficaz no suporte ao pensamento e na percepção humana, bem como na promoção da produtividade do próprio processo (DIEHL, 2002).

Atualmente, houve um aumento no uso da visualização aplicada ao ER (REDDIVARI et al., 2014), bem como nas visualizações na mineração de texto

(KUCHER, KERREN, 2015). No entanto, a ER possui poucos trabalhos interessados na visualização em apoio à mineração de texto.

Esta dissertação tem por objetivo prover um mecanismo para auxiliar ao engenheiro de requisitos, direcionado na divulgação de pesquisas que usam visualização associadas à mineração. Essa divulgação, através de um software, foi estruturada de forma a possibilitar uma busca na qual se associa tipo de mineração com possibilidades de visualização. Assim é viável acelerar a aquisição de conhecimento em tarefas de extração de requisitos por meio de mineração.

1.4 Definição do problema

De acordo com o acima mencionado, a ER precisa de mineração que, para ser facilmente entendida deve contar com o apoio de estratégias de visualização. Essas estratégias de visualização impactam no processo de identificação de informações perceptíveis e localização de relacionamentos implícitos. A seleção adequada da visualização pode tornar o trabalho do elicitador, que usa essa técnica para análise de documentos, mais rápido e bem-sucedido, agilizando a compreensão dos seus resultados. Para isso, no decorrer da pesquisa responder-se-á a seguinte pergunta:

Como melhorar o processo de obtenção de conhecimento sobre visualização de texto minerado na Engenharia de Requisitos?

Como objetivos específicos para atingir a pergunta foram estabelecidos:

- i. Definição de estratégia de classificação das visualizações de texto minerado para Engenharia de Requisitos.
- ii. Criação de uma base de conhecimento que agrupe e organize as visualizações da mineração de texto para Engenharia de Requisitos.

Este trabalho permitirá potencializar o reuso de visualizações que mostram resultados de mineração em tarefas de elicitación, disponibilizando uma estratégia de consulta ou classificação através de uma base de conhecimento. Pretende-se,

com isso, acelerar o processo de obtenção de conhecimento com a finalidade de criar requisitos.

1.5 Organização do documento

Para elucidar os resultados, o trabalho foi estruturado em 5 capítulos além desta Introdução. No capítulo 2 realiza-se o estudo dos dados provenientes da mineração, e as opções de visualização que propõe a literatura. O capítulo 3 caracteriza os relacionamentos entre as tarefas analíticas de mineração, e os tipos de visualização como parte da estratégia para organizar o conhecimento atual. O capítulo 4 detalha a criação da base de conhecimento para organizar os resultados atuais da busca. Também descreve a estratégia para o seu enriquecimento ao longo do tempo (evolução). O capítulo 5 trata da avaliação que foi conduzida a medir a relação qualidade-tempo na obtenção de conhecimento com o uso da estratégia. Por último (capítulo 6), é descrita a conclusão, as contribuições e os trabalhos futuros.

1.6 Sumário

Descrevemos a motivação do trabalho, derivada do uso da mineração de texto na elicitação de requisitos. Tratamos do potencial das técnicas de visualização como facilitadoras do entendimento dos resultados da mineração. Nesse contexto, o foco da pesquisa é facilitar o acesso às informações que associam técnicas de mineração com técnicas de visualização. Finalizamos com o detalhamento das ações para desenvolver a investigação.

2 Revisão de literatura

Neste capítulo realizou-se uma revisão para conhecer o estado da arte das visualizações de texto minerado. Com as conclusões reveladas da revisão, definiu-se as tarefas de mineração que serão usadas para ER e também as categorias de visualização relacionadas.

2.1 Visualização da mineração de texto para ER

A revisão da literatura foi executada em duas etapas. A primeira respondeu a uma pergunta: como as informações de mineração de texto são apresentadas? O objetivo principal era conhecer os diferentes tipos de visualizações.

Partindo da ideia de revisão de literatura foram analisados trabalhos disponíveis na plataforma que integra desenvolvimentos recentes e notícias sobre áreas e aspectos da visualização de informações. O recurso consultado foi: “*Survey of Text Visualization Techniques*” (KUCHER, KERREN, 2015), do Grupo: Informação e Visualização de Software (ISOVIS) da Universidade de Linnaeus, Växjö, Suécia. É uma plataforma que mostra textos relacionados às visualizações de minerações. Baseia-se em uma taxonomia resultante dos refinamentos que ocorreram durante a categorização das entradas, ou seja, a escolha dos elementos da categoria específica é motivada pelos dados enviados pelos próprios atores (KUCHER, KERREN, 2015). A seguir, são explicados cada um dos elementos da taxonomia (KUCHER, KERREN, 2015) que nos ajudarão na seleção e filtragem do total de artigos a serem analisados:

- *Analytic Tasks*: dedicada a tarefas analíticas que são essenciais para as principais metas de análise ao empregar uma técnica de visualização de texto.
- *Visualization Tasks*: seção que descreve as tarefas de representação e interação que são suportadas pelas técnicas de visualização de texto.

- *Domain*: esta categoria descreve os domínios específicos para os quais uma técnica foi desenvolvida.
- *Data and Visualization*: categoria para a relação à fonte de dados e propriedades de dados.

Nesta última seção da taxonomia (*Data and Visualization*) é dividida em várias subcategorias das quais excluiremos aquelas representações em 3D, já que não serão foco de atenção neste trabalho. Como fontes de dados, somos inclinados apenas a *Document*, já que eles formam uma grande parte das informações coletadas na elicitación (TAN, 1999). Com essa seleção, garantimos conhecer o comportamento das visualizações sem aprofundarmos no tratamento de grandes quantidades de textos. A filtragem é baseada em uma operação lógica OR, portanto as classificações que foram excluídas, podem aparecer nos artigos mostrados. Resultaram 78 os artigos para análise (Ver Apêndice I) procurados a partir do ano 2000. Sua análise nos permitiu conhecer as diferentes tarefas que são usadas na mineração de texto e como os seus resultados são apresentados.

Da leitura dos resumos dos artigos encontrados chamou a atenção que muitas das visualizações no período procurado foram evidenciadas para a análise das emoções [ID2, ID3, ID4, ID5, ID8, ID17, ID18, ID26, ID39, ID41, ID54, ID60, ID71, ID75, ID78]. Outro destaque está no uso de atributos de fonte [ID14, ID22, ID25, ID33, ID37, ID65, ID68] junto com diferentes tipos de designs de nuvem de palavras [ID1, ID27, ID36, ID55]. Os mapas de documentos [ID38, ID50, ID53, ID56, ID57, ID59, ID63, ID74, ID77] da mesma forma, com grande uso. Foram vistas representações de semelhança [ID46], frases e termos relevantes [ID6, ID7, ID15, ID16, ID21, ID28, ID29, ID45, ID58], assim como, partes relevantes [ID51, ID52, ID73] de textos. Também é exibido grandes conjuntos de documentos indicando a prevalência de tópico nos corpora [ID31, ID47, ID48, ID64, ID72] ou usam técnicas de agrupamento [ID35, ID67, ID70] para melhorar o entendimento das informações analisadas. Formulários visuais são mostrados para comparar textos [ID11, ID12, ID19, ID30], apresentar resumos [ID61, ID69], além de ajudar a melhorar a experiência de pesquisa [ID24, ID44] do usuário.

Nesta ferramenta, não foram encontrados artigos com aporte para a ER. Se bem essas tarefas podem ser usadas na elicitación de requisitos, nosso objetivo em um

segundo passo, foi esclarecer de acordo com a literatura, após de uma mineração como são apresentadas as informações para ER.

Foi adotada uma estratégia semelhante à executada em uma Revisão Sistemática da Literatura (RSL) de 2016 (ABAD, NOAEEN, RUHE, 2016). Os atores formaram uma *string* de busca com as palavras mais frequentes relacionadas à visualização extraída de uma busca piloto, feita manualmente. Essa *string* elaborada foi aplicada ao Google Scholar e em outros bancos de dados eletrônicos, como: Biblioteca Digital ACM, Science Direct, Springer Link e IEEE Xplore. O foco dessa busca é o uso da UML como técnica de visualização na ER, também evidenciada em DUARTE et al. (2012). Eles concluem que a visualização para esta área tem sido pouco explorada apesar de sua necessidade crescente com o aumento da informação na web (ABAD, NOAEEN, RUHE, 2016). Vale ressaltar que essa RSL se limitou à visualização na ER, sem levar em conta que as informações a serem exibidas estejam mineradas, o que constitui nosso objetivo.

Partindo da ideia de ABAD, NOAEEN e RUHE (2016) a busca bibliográfica feita neste trabalho foi concentrada nas visualizações daquela mineração de textos orientada à ER. O escopo de busca foi reduzido para os espaços com publicações nesta área, Entre eles: IEEE International Requirements Engineering Conference (RE), Requirements Engineering Journal, Workshop em Engenharia de Requisitos, Transactions on Software Engineering and Methodology (TOSEM), IEEE Transactions on Software Engineering (TSE), International Conference on Conceptual Modeling (ER), Brazilian Symposium on Software Engineering, Brazilian Symposium on Computing System Engineering (SBESC), Brazilian Symposium on Software Quality (SBQS) e International Conference on Software Engineering (ICSE). Entre eles foram lidos cada um dos títulos, de cada edição, entre os anos disponíveis. O objetivo perseguido foi não excluir pesquisas que possam ser de interesse. Foram obtidos 89 artigos para uma lista inicial resultante da leitura dos títulos. Como próximo passo foi ler o resumo deles, procurando vínculo entre visualização e mineração na ER. Muitas destes artigos foram descartados, porque abordaram o problema parcialmente; falaram da visualização na ER sem mineração ou vice-versa. Um total de 12 pesquisas foram as resultantes (artigos finais), onde essas três áreas estavam vinculadas (Ver Apêndice II). A

técnica *snowballing* foi aplicada a esses artigos encontrados, uma vez que, foram mencionadas novas referências de interesse e com contribuições de visualizações para mineração em ER, a partir dessa análise foram adicionados mais dois artigos.

Concluindo a busca bibliográfica, podemos concordar com ABAD, NOAEEN e RUHE (2016) em relação à pequena quantidade de pesquisas evidenciando a utilização da visualização para ER com ênfase em aquelas com uso da mineração de texto. Os resultados obtidos, ajudam a revelar que a mineração de texto é aplicada com objetivos diferentes; por exemplo, associar dependências entre requisitos [A6], categorizar requisitos em atributos de qualidade [A8] ou para representar modelos de casos de uso [A1], entre outros.

A leitura completa dos artigos, mostra que nem sempre é possível chegar num relacionamento entre atividades de mineração e suas representações, com a leitura de resumo; por exemplo de [A5], não tinha tarefa de mineração descrita contudo e falaram de grandes quantidades de dados para serem processados. Da mesma forma [A11], descreve tarefas de *Machine Learning* mas com objetivo probabilístico.

As restantes pesquisas revelaram o uso de técnicas de Agrupamento (*Clustering*) para conseguir os seus objetivos [A2, A3, A4, A7, A8] como: melhorar a exploração dos requisitos [A2], agrupar conceitos na engenharia reversa para gerar artefato [A3] e refinar a rastreabilidade [A4, A7, A8]. Outra técnica como Realçado (*Highlighted*) é empregada para destacar trechos de textos com informações relacionada a requisitos [A9] ou termos relevantes para a compreensão de domínio de aplicação [A12]. São evidenciadas pesquisas q usam cálculo de Similaridade (*Similarity*) para agrupar requisitos de acordo com sua descrição [A2] além da detecção de dependência para rastreabilidade a partir de matrizes [A6]. O Rotulado (*Labeled*), é uma tarefa que não é usada solitariamente. Para algumas pesquisas [A3, A8, A12, A13], ela é exibida conjuntamente com ressaltado de um termo o para agrupar esses termos de acordo com um critério; salvo em [A10] quando é usado para obter classificação do requisito, em [A1] para facilitar a transformação para um diagrama de caso de uso e [A14] para obter modelos de especificação mais refinados. Para acentuar o resultado encontrado, uma ferramenta chamou a atenção [A12], ela tem como objetivo compreender melhor o domínio de aplicação com extração dinâmica e etiquetado de conhecimento relevante de requisitos.

Aparentemente permite tarefas de Realçado e Rotulado, mas no estudo comprovou-se que, do mesmo modo, resume trechos de texto para expor os fragmentos relevantes. Incluindo, a partir da análise, o Resumo (*Sumarization*) como outra tarefa executada de mineração. Deve-se salientar que em muitos casos essas tarefas foram combinadas, como no caso de [A2, A3, A8, A12, A13] que adotam mais de uma técnica.

Com o intuito de entender a atribuição dessas tarefas a cada artigo, foi necessário estabelecer certas definições listadas na seção seguinte.

2.2 Tarefas analíticas

As tarefas analíticas explicadas na Seção 2.1 resultantes da busca bibliográfica e executadas pela mineração de texto especificamente para ER são as seguintes: destaque do texto (*Realçado*), aplicação de técnicas de agrupamento por domínio ou requisitos (*Agrupamento*) ou pelo grau de similaridade (*Similaridade*); rotulagem de conhecimento relevante (*Rotulado*) e resumo (*Resumo*) onde *Realçado*, *Rotulado* e o *Agrupamento* são os mais utilizados.

É uma necessidade para esta pesquisa definir o comportamento das tarefas encontradas, com o objetivo de entender quando são usadas. Partindo da definição da palavra na linguística, foi feita a associação com a função pela que foram empregadas.

- i. *Agrupamento*: é atribuído a aquelas pesquisas que agrupam, unem e associam informações. Como é o caso de requisitos do mesmo tipo ou documentos do mesmo domínio.
- ii. *Realçado*: é outorgado quando se consegue destacar com cores, tamanho e outros recursos de estilo. Exemplificando ao realçar seções dentro de um texto que podem ser relevantes para construção de requisitos com cor.
- iii. *Resumo*: sintetizar ideias chaves com frases. Como na extração de trechos relevantes nas descrições.
- iv. *Rotulado*: designação de característica definidora em poucas palavras. Por exemplo os atributos de qualidade de requisitos, ou a identificação de entidades para a construção de modelos.

- v. Similaridade: atribuído quando se procura informações da mesma natureza, semelhantes. Ao representar similaridades nas descrições.

Dada essa visão de tarefas analíticas, é substancial retornarmos à ferramenta “*Survey of Text Visualization Techniques*” (KUCHER, KERREN, 2015), com a finalidade de vincular a visão desta pesquisa com a proposta (KUCHER, KERREN, 2015). Na ferramenta as tarefas que precisamos analisar são aquelas da seção Tarefas de Visualização ou *Visualization Task* que descrevem as tarefas de representação executadas para visualizar texto:

- i. *Clustering/Classification/Categorization*: propicia várias tarefas relacionadas à marcação (semi-) automática ou agrupamento de elementos de dados.
- ii. *Overview*: técnicas que usam representações especiais para fornecer uma visão geral e reduzir a complexidade visual.
- iii. *Comparison*: disposição de vários objetos lado a lado ou discrepância de marcação.
- iv. *Navigation/Exploration*: facilitam o processo de navegação no conjunto de dados, enquanto alternamos as representações visuais ou os tipos de dados subjacentes.
- v. *Region of Interest*: esta tarefa indica o realce / sugestão automática de itens / regiões de dados que podem ser do interesse.
- vi. *Uncertainty Tackling*: manipulam e/ou visualizam incertezas nos dados de origem ou processados, bem como incertezas nos cálculos.
- vii. *Monitoring*: esta tarefa está relacionada às técnicas de visualização projetadas para alertar os usuários sobre as alterações nos dados.

Analogamente às tarefas encontradas na busca bibliográfica está presente Agrupamento em *Clustering/Classification/Categorization*, Realçado equivalente a *Region of Interest*, Rotulado e Resumo similar ao resultado de *Overview*, Similaridade parecido a *Comparison*. Não se encontrou evidência do uso para ER de *Monitoring*, *Uncertainty Tackling* e *Navigation/Exploration*.

A aplicabilidade das técnicas que não foram encontradas na busca bibliográfica será analisada de acordo com a sua própria função para tarefas na ER. Como é, por exemplo, visualizar a incerteza nos dados de um domínio (*Uncertainty Tackling*), é

como criar um software instável, saindo do princípio básico: saber o que se deseja antes de construí-lo. Por outro lado, o monitoramento das alterações nos dados (*Monitoring*) é evidenciado no gerenciamento de requisitos, ao longo do processo de software. Nesta pesquisa, o foco está nas tarefas de elicitar requisitos, o que pode ser levado em consideração para a evolução deste trabalho. O mesmo acontece com (*Navigation/Exploration*), já que propõe diferentes perspectivas para ver os próprios dados, mas o nosso interesse inicial está focado diretamente nas tarefas da ER e nem tanto na granularidade dos resultados a mostrar, mas seria uma inclusão interessante. No entanto esta pesquisa pretende apresentar diferentes propostas de visualização para uma determinada tarefa de mineração de acordo com o que existe na literatura.

2.3 Categorias de visualização

Tudo o anteriormente descrito não menciona como são classificadas as visualizações encontradas na RSL executada. Com relação a essa necessidade, foi possível consultar a ferramenta “*From Data to Viz*” (HEALY, [s.d.]), resultado de um projeto desenvolvido por franceses, que apresenta a classificação dos tipos de visualizações com base no formato de dados de entrada o que é mostrado numa árvore de decisão.

Usaremos a nomenclatura original (em inglês) para preservar seu significado. A razão principal dessa decisão tem a ver com a insuficiente bibliografia referenciando dita nomenclatura em português. Forçar uma tradução pode atentar contra a compreensão dos termos e do contexto.

A consulta apresenta os tipos de categoria de acordo com o gráfico de visualização (HEALY, [s.d.]), elas são:

- *Distribution*: evidencia comportamento da distribuição de uma variável ou de vários grupos.



Figura 1 Categoria de Visualização: *Distribution*.

- *Correlation*: permitem apresentar a comparação entre variáveis. Relaciona par de variáveis numéricas e pode acrescentar mais uma variável para representar o tamanho e importância.

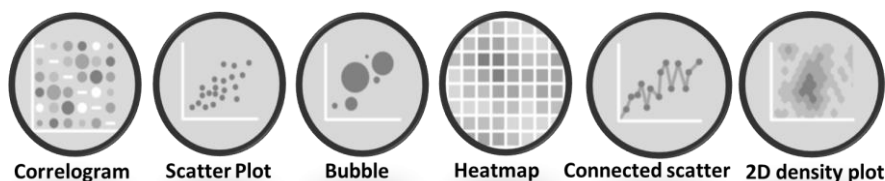


Figura 2 Categoria de Visualização: *Correlation*.

- *Ranking*: apresenta gráficos como *Wordcloud* que exibe uma lista de palavras, com a importância de cada uma mostrada com tamanho ou cor da fonte. Este formato é útil para perceber rapidamente os termos mais importantes. Funciona bem para revelar o essencial.

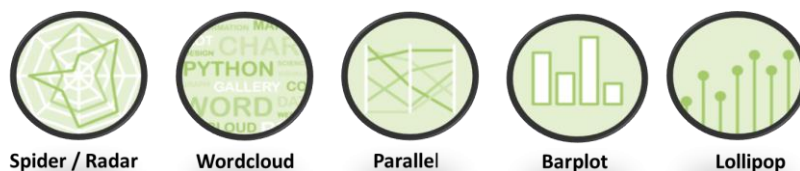


Figura 3 Categoria de Visualização: *Ranking*.

- *Part of Whole*: visualiza arquiteturas hierárquicas. Os gráficos têm a vantagem de fazer uso eficiente do espaço, o que os torna úteis para representar uma grande quantidade de dados. Pode ser recomendado para criar figuras interativas.

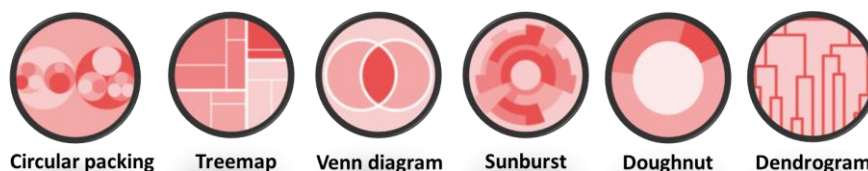


Figura 4 Categoria de Visualização: *Part of Whole*.

- *Evolution*: apresenta a evolução das variáveis numéricas. Além de versões empilhada de gráficos de áreas para comparar vários grupos.



Figura 5 Categoria de Visualização: *Evolution*.

- *Map*: mostra gráficos baseados em coordenadas geográficas.

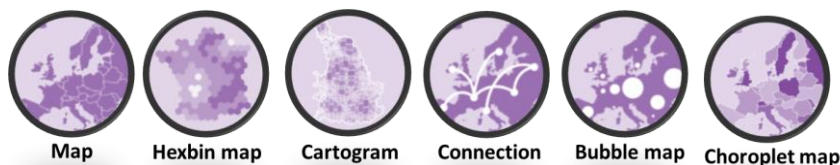


Figura 6 Categoria de Visualização: *Map*.

- *Flow*: permite exibir conjunto de nós interconectados propiciado exibir fluxos.



Figura 7 Categoria de Visualização: *Flow*.

De acordo com as categorias propostas pela ferramenta “*From Data to Viz*” (HEALY, [s.d.]), uma filtragem foi realizada para manter apenas aqueles que podem contribuir para as tarefas de elicitación na ER. Nesse sentido a categoria de *Distribution* é descartada aqui porque está focada em conhecer a distribuição de uma única variável. Foi desconsiderada, também, *Evolution* porque seu objetivo é apresentar a tendência dos dados em intervalos de tempo. *Map* também foi excluído, pois está baseado em coordenadas geográficas.

2.4 Sumário

Nesta seção foi relatado o processo seguido para obter as informações conceituais da pesquisa. Um resumo da busca bibliográfica com os principais trabalhos analisados foi descrito. Foram adotadas posições conceituais das áreas de mineração de texto e da visualização, para serem atribuídos ao longo do trabalho.

3 Correlação das tarefas de mineração de texto com categorias de visualização para ER

Este capítulo descreve a correlação das tarefas de mineração de texto consideradas para ER, com as categorias de visualização. Além de revelar as condições para implementação dos gráficos dessas categorias.

3.1 Componentes

Dispor de ferramentas que ajudem o dia a dia do engenheiro é desejável devido à quantidade de informação a processar e à competitividade do mercado. Podemos observar na Figura 8 parte da estratégia pretendida para integrar a mineração e a visualização na ER. A estratégia persegue a integração dada pela

entrada da tarefa analítica para obter a sugestão de visualização como saída, onde nesse processo incidem três aspectos fundamentais a considerar:

- i) Tipos de tarefas analíticas: onde é definida a classificação da tarefa de mineração a executar.
- ii) Tipos de dados: é parte do resultado da mineração que ajuda nas possibilidades de apresentar o resultado final desejado.
- iii) Categorias de visualização: é a seleção da alternativa de visualização.

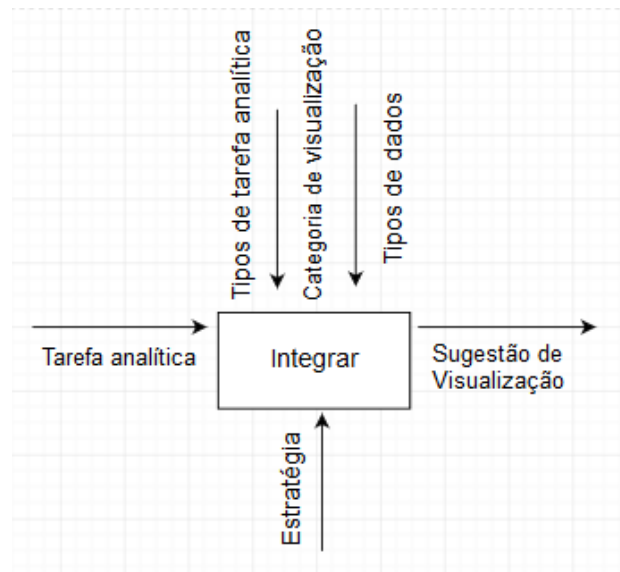


Figura 8 Estratégia para Integrar Mineração de texto com Visualização.

A análise detalhada da Figura 8 considera a Tarefa analítica com a qual o texto será minerado, o ponto inicial. No processo de **Integrar** é primordial estabelecer níveis de execução para seu entendimento. Primeiramente está o nível de *Identificar* o Tipo de tarefa analítica, seguidamente *Separar*, de acordo com o Tipo de dado gerado da tarefa analítica e com isso *Associar* uma categoria de visualização apropriada, conforme o refletido na Figura 9. Para o Tipo de tarefa de mineração serão as definições da Seção 2.2, usadas na ER: Agrupamento, Realçado, Resumo, Rotulado ou Similaridade. A execução delas, contempla a transformação dos dados

brutos em saídas numéricas ou categóricas. Essa transformação é mais um fator que permitirá escolher uma categoria de visualização apropriada.

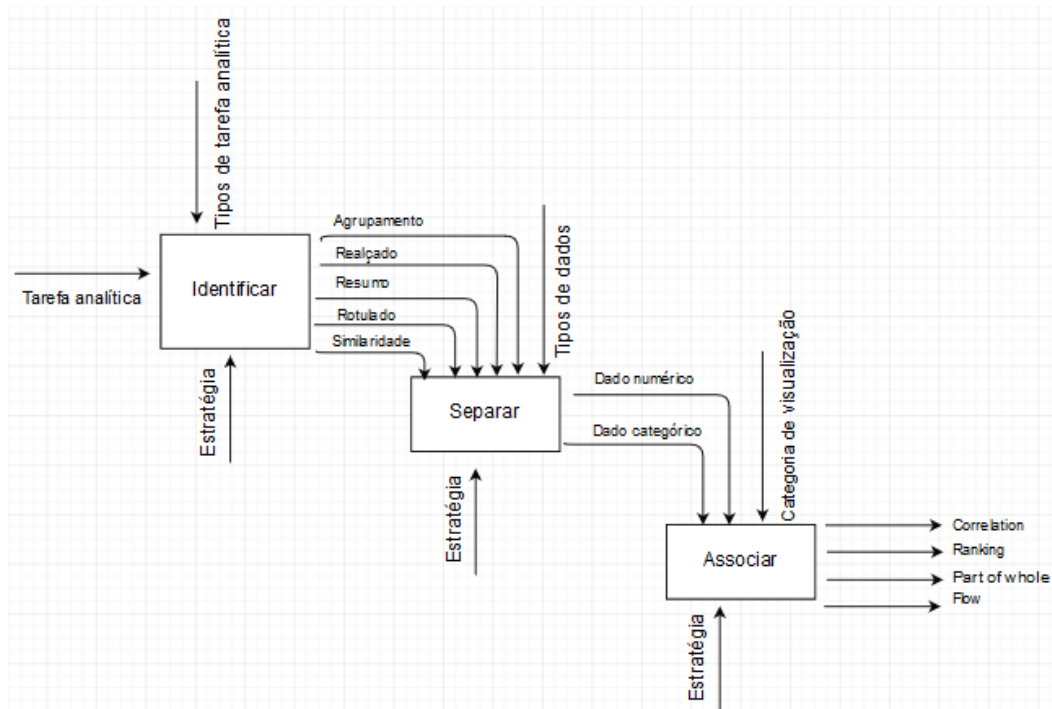


Figura 9 Modelo da inter-relação dos componentes.

Entre os dados categóricos podem-se ser evidenciados: as palavras, sentenças, parágrafos, documentos e corpora, ou seja, texto, resultantes após a mineração, porém, passíveis de representação (BRATH; BANISSI, 2016). A representação deles será explicada na Seção 3.1.1. Por outro lado, os dados numéricos referem-se aos conjuntos de números reais como, por exemplo, o preço em media de 10.000 apartamentos do Airbnb na Riviera Francesa na França (HEALY, [s.d.]).

Os fatores a serem levados em consideração para a seleção de uma visualização estão expressos na Tabela 1 em forma de resumo (HEALY, [s.d.]), começando com as Categorias que foram evidenciadas para a ER na seção 2.3, seguido pelo Tipo de Gráfico como uma alternativa para cada categoria. Finalmente, são exibidos o tipo de dados e o número de variáveis necessárias para implementar este gráfico no código.

Exemplificando um pouco melhor, imaginemos que é desejável usar *Wordcloud* (conhecida além por nuvem de etiquetas ou lista ponderada) para visualizar um conjunto de dados. Os requisitos que devem cumprir esse conjunto de dados são: 1) que seja uma única variável, traduzido a coluna única de valores e 2) que os valores

dessa coluna sejam categorias, o seja: palavras, frases, etc., refletido como Tipo de dado.

Tabela 1 Tipo de dados por categorias.

CATEGORIA	TIPO DE GRÁFICO	TIPO DE DADO	NO. VARIÁVEIS	
<i>Correlation</i>	Scatter	Numérico	Duas variáveis (Não ordenadas)	
	Bubble Plot	Numérico	Três variáveis (Não ordenadas)	
	Correlogram	Numérico	Várias (Não ordenadas)	
	Heatmap	Numérico	Várias (Não ordenadas)	
	Connected Scatter Plot		Categórico	Uma variável/Subgrupo/Adjacência
			Numérico	Duas variáveis (Ordenadas)
<i>Ranking</i>	Barplot	Categórico	Uma variável/Nested/Subgrupo	
	Spider / Radar	Categórico	Subgrupo	
	Wordcloud	Categórico	Uma variável	
	Parallel	Categórico	Subgrupo	
	Lollipop	Categórico	Uma variável/Subgrupo	
<i>Part of whole</i>	Dendrogram	Numérico	Várias (No ordenadas)	
		Categórico	Nested	
	Treemap	Categórico	Uma variável/Nested	
	Venn diagram	Categórico	Duas listas de categorias	
	Doughnut	Categórico	Uma variável	
	Circular packing	Categórico	Uma variável/Nested	
	Sunburst	Categórico	Nested	
<i>Flow</i>	Chord diagram	Categórico	Adjacência	
	Network	Categórico	Adjacência	
	Sankey		Categórico	Subgrupo/Adjacência
			Categórico	Adjacência
	Arc diagram	Categórico	Adjacência	

É evidenciado na Tabela 1, como a categoria *Ranking* e *Flow* propõe cada visualização dependendo dos dados categóricos. Outra perspectiva sob os mesmos tipos de dados, na categoria *Part of whole*, onde quase todas as suas representações precisam deles. Por outro lado, a categoria *Correlation* possui um número maior de representações que permitem dados numéricos em comparação com os dados categóricos.

Na próxima seção trata-se o comportamento e escopo dos dados categóricos e como eles podem aportar para o processo da mineração na ER.

3.1.1. Dados das tarefas analíticas

Apenas analisamos os dados categóricos, já que não encontramos o tipo de dado numérico na literatura de mineração de texto em ER.

Em primeiro lugar é abordada a palavra, sendo a menor estrutura da sentença. A palavra pode ser tratada como etiqueta tanto para diagrama, quanto classificação do requisito, como é o caso de: *Segurança, Controle de Acesso*. Consequentemente será definida como:

Palavra: componente de uma sentença que tem significado próprio

(CALVO, 1982).

De outro ponto de vista, as frases ou sentenças são informações relacionadas ao requisito, título ou palavra-chave no contexto. Ex. *“O aplicativo faz chamadas criando um hash MD5 de uma chave pública e privada, bem como um carimbo de data e hora para autenticar a chamada.”* Em vista disso, pode ser definida como:

Sentença: Composição de palavras que transmitem uma ideia

(“Dicio”, 2019b).

Descrições de informações referentes ao contexto e resumo de poucas sentenças podem ser representados pelo parágrafo. Ex (LATHAM, 2017). *“O serviço permitirá que os usuários joguem um jogo no qual marcam as letras das músicas com palavras-chave emotivas. Em seguida, esses dados podem ser usados para coisas interessantes, incluindo sugestões de músicas para os usuários. À medida que os dados sobre marcação são coletados, os usuários que responderem com as etiquetas mais populares serão recompensados de alguma forma. Inicialmente, o sistema receberá um número de artistas, álbuns, músicas e letras de músicas do Last.fm e serviços de letras musicais”.* Com esta visão pode-se definir:

Parágrafo: Conjunto de sentenças com uma ou várias ideias

(“Dicio”, 2019a).

Já os documentos e corpora são representações mais amplas que incluem vários parágrafos com um objetivo, tópicos e ideias variadas. Ex (“Mac Morpho”, [s.d.]). *Brazilian Portuguese news text with part-of-speech tags*, disponível: <https://www.kaggle.com/nltkdata/mac-morpho>

Os conceitos acima dão uma ideia dos resultados esperados na mineração de texto. Embora estas saídas possam ser convertidas em números, de acordo com qualquer método de mineração, nosso alvo consiste em identificar como são exibidos.

De maneira geral os dados categóricos podem ser evidenciados nas saídas de um processo de mineração de texto. A análise seguinte salientará o seu comportamento com a representação visual, apontando finalmente ao relacionamento entre estas tarefas de mineração e as categorias de visualização para a ER.

3.1.2. Relação entre tarefa analítica e categoria de visualização

No modelo da Figura 9, a correlação entre tarefas e categorias é identificada pelo tipo de dados enviados e recebidos. É uma correlação que será analisada a seguir, com base em dados categóricos como entrada para visualização.

Realçado

Tomando como exemplo a tarefa analítica Realçado, podemos distinguir o cenários do uso de **parágrafo** como dado categórico. A Figura 10, exemplifica como é potenciado o uso cor de fundo para realçar. Destaca-se o texto mais relevante na descrição de um componente específico: “Unidade de bordo” (OBU).

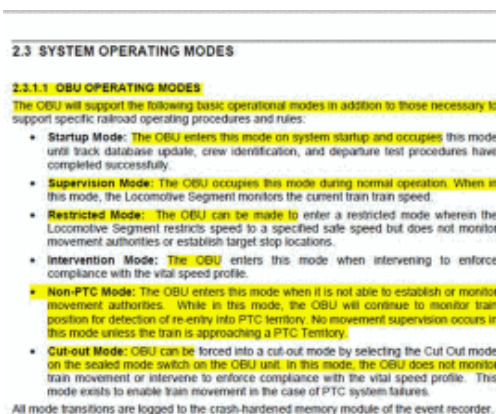


Figura 10 Realçado em documento. Ref. Apêndice II No. [9].

Resumo

Por outro lado, o Resumo, embora não seja explicitamente evidenciado para ER, foi adotado como outras das tarefas analíticas da mineração, uma vez que serve como parte da execução de tarefas como Agrupamento e Rotulado. A definição da Seção 2.2 esclarece a possibilidade de ser empregada para sentenças, parágrafos ou

documentos; excetuando a palavra já que a concepção da tarefa é diminuir o dado inicial. A Figura 11 revela um resumo em texto livre, que conta com menor quantidade de sentenças onde, a partir de um artigo original de 26 sentenças, o resumo de documentos ficou com seis.

As Ghip told SPIEGEL ONLINE, it all started when he was confronted with a few small, practical problems. So Ghip put a cat door on the people door and solved his first problem.

The second problem arose when Ghip figured out that the neighbors' cats were sliding in and out of the cat door for free meals because, as he says on his Web site, they "learned that the food my wife buys is better than what they get from (their own) owners."

A servo is a small mechanical devise that accepts signals and is commonly found in radio-controlled airplanes, puppets and robots.

The RFID reader sends a message to the laptop and the software on the laptop then sends a message to the servo controller, which then tells the servo whether or not to open the door using power from the USB connection to the computer.

"Anyone with some tools and time on their hands can do it," Ghip told SPIEGEL ONLINE.

Figura 11 Resumo de documento. Ref. Apêndice I No. [61].

A falta de representações para o Resumo é mais um problema detectado por esta pesquisa, além de que a visualização existente não possui correspondência na galeria de categorias analisadas (Veja Seção 2.3).

Agrupamento

A tarefa seguinte de análise é o Agrupamento de temas ou características semelhantes entre os dados exemplificado na Figura 12 que ilustra o agrupamento para frases.

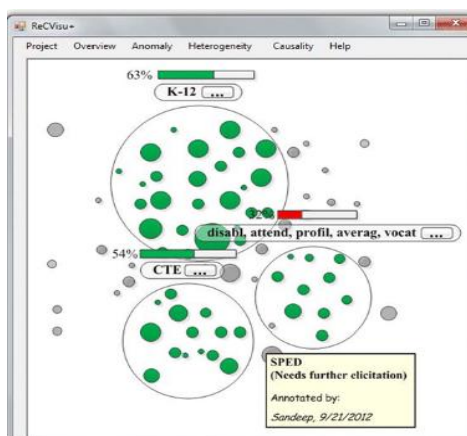


Figura 12 Agrupamento de frases. Ref. Apêndice II No. [2].

Cabe ressaltar que o agrupamento é principalmente usado com *Circular Packing*, que pertence à categoria *Part of Whole* e *Bubble*, que pertence a categoria de *Correlation*. O primeiro deles, semelhante à *Treemap*, permite visualizar uma arquitetura hierárquica onde cada nó da árvore é representado como um círculo e seus subnós são representados como círculos dentro dele. O *Bubble Plot* é um gráfico de dispersão onde uma terceira dimensão é adicionada: o valor de uma variável numérica adicional é representado pelo tamanho dos pontos. Se continuarmos com a ideia da representação de dados categóricos, os dados para esse último gráfico mencionado serão alvo de conversão para conseguir sua utilização.

Acrescentando a abordagem explorada, também foram avaliados cada uns dos artigos resultantes da busca focada à ER. O exemplo do ReCVisu (REDDIVARI, CHEN, NIU, 2012), que é uma ferramenta que ajuda à exploração de requisitos. Entre suas tarefas analíticas estão o agrupamento (Agrupamento) e rotulado (Rotulado). O resultado visual de ReCVisu é mostrado com *Circular Packing*. A identificação dos níveis de abstração em cada uns dos agrupamentos feitos é a intensão da visualização nela.

Similaridade

A próxima categoria a analisar é a Similaridade que está refletida na proposta da Figura 13 com o desenvolvimento de uma matriz de rastreabilidade (Referenciado do artigo no. 6 do Apêndice II). As descrições de cada requisito, sentenças, são convertidas em dado numérico para calcular a semelhança entre eles através de uma heurística. O gráfico proposto na Figura 13, pode ser homologado de acordo com as representação das categorias propostas na Seção 2.3 com o uso de *Heatmap* da categoria *Correlation*, porque exhibe uma visão geral dos dados e a intensidade da cor permitirá mostrar pontos semelhantes a destacar (HEALY, [s.d.]).

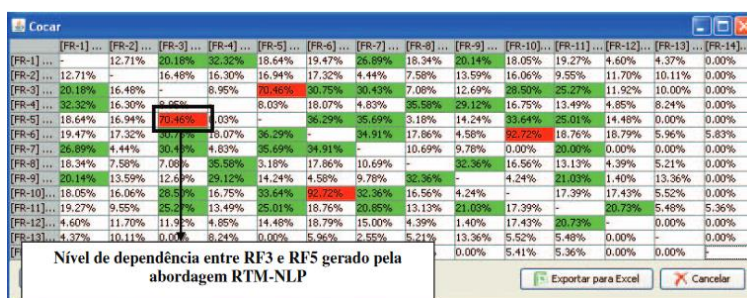


Figura 13 Similaridade em frases. Ref. Apêndice II No. [6].

Rotulado

Algumas das tarefas analíticas listadas exibem combinações entre si para cumprir o objetivo traçado. O Rotulado é um dos que dificilmente é exibido como tarefa atômica. Embora tenha sido difícil encontrar uma pesquisa para exemplificar, a Figura 14 apresenta um *Softgoal Independence Graph* (SIG) como resultado da engenharia reversa proposta para atributos de segurança.

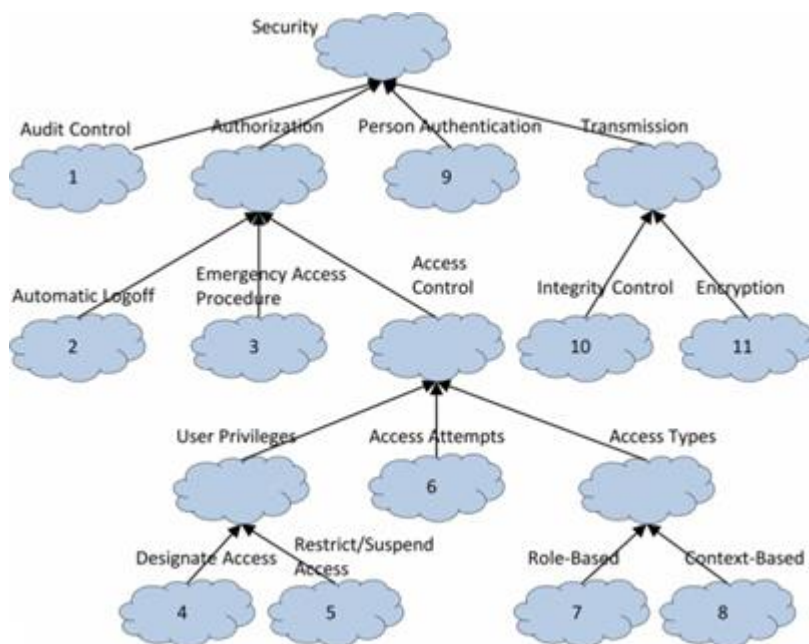


Figura 14 Rotulado em sentenças. Ref. Apêndice II No. [8].

A Figura 14 expõe uma hierarquia de conceitos em que foi definida a segurança como atributo de qualidade. Uma forma de árvore é exibida seguindo o padrão das gráficas da categoria de *Flow* ou *Dendrogram* da categoria de *Part of Whole*. O *Dendrogram* pode ser feito com 2 tipos de conjunto de dados segundo a Tabela 1: i) uma matriz numérica que permite calcular a distância entre indivíduos; ii) um conjunto de dados hierárquico em que o relacionamento entre entidades é fornecido diretamente.

É conveniente empregar a categoria de visualização *Flow*, onde os nós ganham protagonismo para conhecer informações curtas. As setas podem representar vínculos entre esses nós e podem ter pesos ou prioridades. É uma categoria de visualização que torna compreensível o uso do Rotulado.

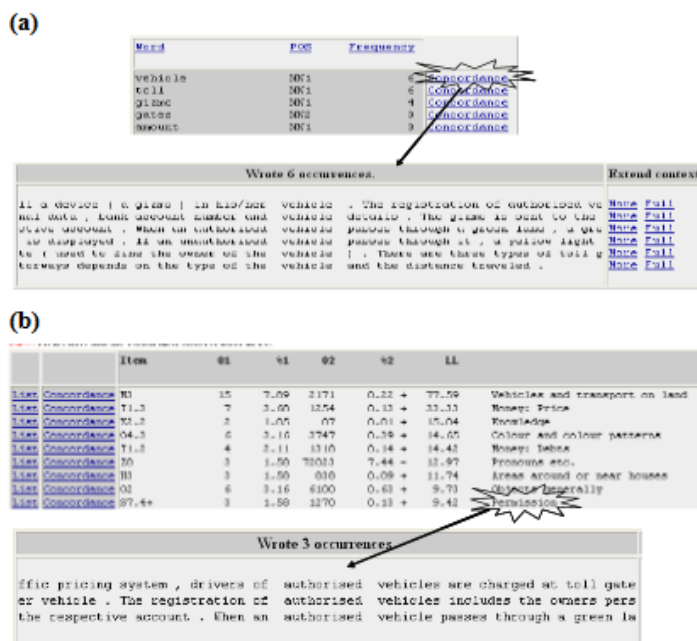


Figura 15 Rotulado de documento. Ref. Apêndice II No. [14].

Sob outra perspectiva, a Figura 15(a) apresenta uma lista de alguns substantivos identificados com a ferramenta EA-Miner (Ref. Apêndice II No. [14]). A Figura 15(b) exibe como a análise semântica agrupa umas palavras com outras do seu mesmo campo semântico para a identificação de requisitos, ex.: identificou do mesmo campo semântico as palavras “Autorizado” e “Permissão”, rotulando assim as sentenças para extrair os requisitos não funcionais.

Conforme o exposto na Figura 15, identificar qualquer gráfico de visualização para essa representação é considerado um tanto difícil, já que a ferramenta da Figura 15 gera lista de palavras e as categorias de visualização analisadas (Veja Seção 2.3) não tem gráficos que possam ser semelhantes.

3.2 Sumário

São explicadas as partes mais relevantes da estratégia, orientada ao processo do relacionamento entre tarefa analítica executada e a categoria de visualização, onde as representações dos dados de saídas foram exploradas com objetivo de

atingir essa associação. Cada anotação dos dados categóricos em conjunto com seu relacionamento com a categoria de visualização permitiu entender que:

- é identificado uma brecha com inexistência de representações das palavras e corpus, além de ressaltado para documentos,
- igualmente existe uma lacuna para apresentar opções de Resumo,
- as categorias de visualização propostas (Seção 2.3) não consideram o uso de cores de fundo para as visualizações a serem usadas na ER, e
- existem gráficos das categorias de visualização que não tem sido explorados na ER.

As poucas visualizações para corpus devem-se principalmente à possibilidade de transformá-los em conjunto de parágrafos, sentenças ou palavras e manipulá-los como tal. Dificilmente são encontrados apresentando Similaridades, Realçado ou Agrupamentos.

Para palavras, por outro lado, têm sido pouco exploradas tarefas como Similaridade, Rotulado ou Resumo. Perde-se a essência o fato de resumir uma palavra ou encontrar similaridade com outra palavra. Embora fosse evidente a saída da palavra nas tarefas analíticas de Realçado para a literatura em geral (KUCHER, KERREN, 2015), foi difícil encontrar o uso para os artigos analisados focados na ER.

Tabela 2 Associação entre Tarefa e Categoria.

CATEGORIAS DE VISUALIZAÇÃO	CORRELATION						RANKING				PART OF WHOLE					FLOW						
	Scatter	Bubble Plot	Correlogram	Heatmap	Correlogram	Connected Scatter Plot	Barplot	Spider / Radar	Wordcloud	Parallel	Lollipop	Dendrogram	Treemap	Venn diagram	Doughnut	Circular packing	Sunburst	Chord diagram	Network	Sankey	Arc diagram	Edge bundling
TAREFAS ANALÍTICAS																						
AGRUPAMENTO																						
REALÇADO																						
SIMILARIDADE																						
ROTULADO																						

A Tabela 2 mostra o resumo das tarefas explicadas com as visualizações executadas que foram evidenciadas na literatura para a ER. De acordo com isso, é disponibilizado uma relação de uso para ER de tarefa analítica com as possíveis representações.

Pode-se concluir que o resultado de toda tarefa analítica tem diferentes representações visuais sendo nesta estratégia sintetizadas pelo aporte para a ER. Em seguida, como influenciam os dados após da mineração no tipo de visualização. Por fim, conseguir uma relação entre eles de acordo com resultados estudados na literatura. Uma avaliação dos comportamentos gráficos das visualizações foi realizada para contribuir com o uso em tarefas analíticas de mineração.

Foram analisadas as tarefas analíticas importantes para ER e as categorias de visualização com as quais elas estão relacionadas, resumidas na Tabela 2. Além disso, foram identificados os artigos que mostram essa relação com o objetivo que possa ser consultado futuramente e ajudem a entender como ele foi feito.

4 Biblioteca Digital de Apoio

O relacionamento entre as tarefas analíticas e as categorias de visualização para ER obtidas com a análise do Capítulo 3 motivou a criação de uma biblioteca digital. A biblioteca coleta essa associação e apresenta os artigos exemplificando como é visualizada a tarefa de mineração executada. O alvo principal da biblioteca é concentrar e armazenar as principais ideias de pesquisas que implementam mineração de texto e apresentam visualizações para seus resultados, bem como o “*Survey of Text Visualization Techniques*” (KUCHER, KERREN, 2015), neste caso para ER.

4.1 Definição

De acordo com Schwartz (2000) uma Biblioteca Digital pode ser definida como uma coleção eletrônica distribuída que pode compreender praticamente todos os campos do empreendimento humano. Eles vão desde coleções especializadas que se concentram em um tipo de arquivo em uma área temática, até agregações complexas de vários formatos que suportam pesquisas interdisciplinares. Uma Biblioteca Digital permite coletar, armazenar e organizar informações em formato digital, além de disponibilizá-las para busca, recuperação e processamento através de redes de comunicação. Outros aspectos que devem ser destacados são: a preservação a longo prazo da informação e, com ela, o conhecimento.

O objetivo da biblioteca desenvolvida é organizar a inter-relação entre tarefas de mineração e categorias de visualização identificadas para uso na ER, que foi explicada na Seção 3.1.2. Exibe a partir dessa organização as pesquisas que sustentam cada relação.

Esta biblioteca foi desenvolvida com método ágil. De acordo com o Manifesto Ágil, o foco está no desenvolvimento de software funcional, em vez de obter um software bem documentado (MARTIN FOWLER et al., 2001).

Dentre as tecnologias foi utilizada a linguagem Python em versão 3.6, que mantém clareza na sintaxe além de ser dinâmica. O framework para desenvolvimento foi o Django, com o qual aplicações de qualquer complexidade podem ser desenvolvidas. Foram incluídas as bibliotecas: PIL para visualizar imagens, JQuery para trabalhar com Java Script incluindo, Bootstrap como um conjunto de ferramentas de código aberto para apoio ao desenvolvimento da interface gráfica, HTML juntamente com CSS e JavaScript. Para o controle de versão foi usado o Git. Dessa forma, a rastreabilidade dos requisitos e a disponibilidade em tempo real do projeto foi garantida.

4.2 Funcionalidade

Embora não esteja disponível o código fonte da ferramenta web *Survey of Text Visualization Techniques*, foi alvo de estudo para elicitare os requisitos da nova Biblioteca Digital. Foram aplicadas técnicas de engenharia reversa experimentadas sobre a interface gráfica dessa ferramenta conjuntamente com a análise dos artigos resultantes da busca bibliográfica. Algumas observações com relação ao funcionamento e design da ferramenta web *Survey of Text Visualization Techniques* são listadas:

No painel superior:

- Contém um resumo de estatística descritiva de acordo com os artigos guardados pela ferramenta.
- Tem um link a outras ferramentas visuais, *surveys* e fórum da comunidade.
- Permite solicitar pelo usuário sua entrada (inserir artigo), como uma nova contribuição. Nessa opção, são definidos os dados num formulário para gerar uma entrada em formato JSON que, unido à imagem, serão enviados pelo e-mail.

Menu esquerdo

- Dispõe de uma taxonomia obtida a partir dos resultados de refinamentos, ocorridos durante a categorização de entradas, ou seja, a escolha de itens da categoria concreta é motivada pelos dados enviados.

- *Analytic Tasks*: dedicada a tarefas analíticas que são essenciais para as principais metas de análise ao empregar uma técnica de visualização de texto.
- *Visualization Tasks*: seção que descreve as tarefas de representação e interação que são suportadas pelas técnicas de visualização de texto.
- *Domain*: esta categoria descreve os domínios de texto dedicados para os quais uma técnica foi desenvolvida.
- *Data and Visualization*: categoria para a relação entre a fonte de dados e as propriedades deles.
- Permite filtrar por um campo de pesquisa de texto, controle deslizante de ano de publicação, assim como os botões das categorias. As técnicas são atribuídas a conjuntos de abas de categoria e a filtragem é baseada na operação lógica OR.

O software desenvolvido: Biblioteca Digital de Visualizações para Mineração de Texto para Engenharia de Requisitos (*MinViewREDL*) conta com três papéis principais:

- Usuário contribuinte: que são os pesquisadores da comunidade da ER que enviarão artigos novos para enriquecer a base de conhecimento da *MinViewREDL*, onde desenvolvem visualizações para mineração de texto.
- Usuário gerenciador: quem armazenará os novos artigos na *MinViewREDL*.
- Usuário pesquisador: quem usará a *MinViewREDL* para fazer buscas.

As metas elicítadas para a Biblioteca digital são listadas a seguir:

- Permite ao gerenciador inserir tanto os artigos, quanto as suas classificações desde uma interface própria.
- Permite ao pesquisador a busca dos artigos pelo título e autor.
- Permite ao pesquisador a filtragem seguindo operação lógica OR para especificidades das entradas: tarefas e categorias de visualização. Ou seja, que funciona como consulta de agrupamento por tarefas de mineração e categorias de visualização.

- Permite selecionar uma das entradas e exibir os detalhes correspondentes a ela, incluindo a(s) tarefa(s) analítica(s) que desenvolverem nesse artigo e a(s) categoria(s) de visualização para apresentar os resultados.

É importante destacar que durante o ato de inserir novos artigos, as tarefas analíticas de mineração e as categorias de visualização correspondentes são catalogadas no momento de serem armazenadas. Com base nessas informações, filtros de pesquisa podem ser criados.

O detalhamento das metas acima é fornecido por estórias de usuário (Apêndice V).

4.3 Arquitetura da *MinViewREDL*

A arquitetura define um padrão arquitetônico Modelo Visão Controlador. Assim, foram definidas a Camada de Interface de Usuário (Templates), a Camada de Lógica de Negócio (Views) e a Camada de Acesso aos Dados (Models), respaldado pela arquitetura do Django da Figura 16.

A arquitetura de Django permitiu a estrutura organizacional da *MinViewREDL* e, a partir dela, o diagrama de componentes segue a estrutura da Figura 17. Na camada de *Template*, encontra-se os componentes de extensão .HTML, .JS e .CSS, que são encarregados de mostrar o conteúdo para ser consultado pelo usuário. O componente *check.js*, conduz toda a lógica de comprovação e filtrado dos artigos tanto na seleção pelo menu esquerdo quanto pelo painel superior de busca por texto. Os componentes HTML contêm o conteúdo da biblioteca além dos detalhes dos artigos para exibi-los.

A *Views* é a classe principal para gerenciar as solicitações, é invocada a partir da URL, que é o ponto inicial de entrada ao sistema. De acordo com a definição de Django que faz roteamento a partir do URL.

Foi reutilizada a interface de Gerenciador de Django (admin app) para armazenar as entradas dos novos artigos. Além disso, foi incorporada uma função de comprovação que verifica para cada artigo novo, as relações definidas entre tarefas de mineração e a categorias de visualização a serem inseridas e que correspondem com o definido na Tabela 2. Essa relação está explicada na Figura 18.

Além disso, é garantido o gerenciamento de novas entradas tanto de novas categorias de visualização quanto de tarefas de mineração concebida futuramente para ER.

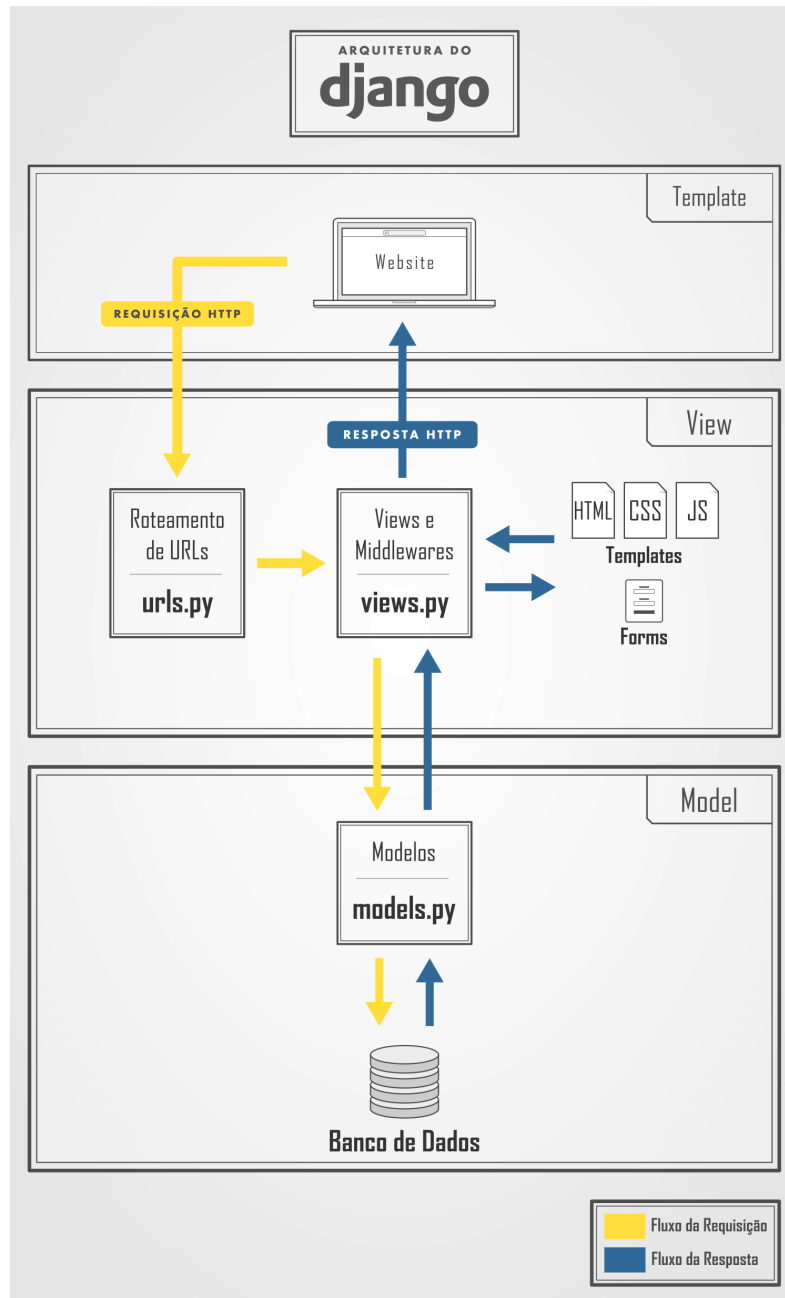


Figura 16 Arquitetura de Django (DJANGO SOFTWARE FOUNDATION, 2019).

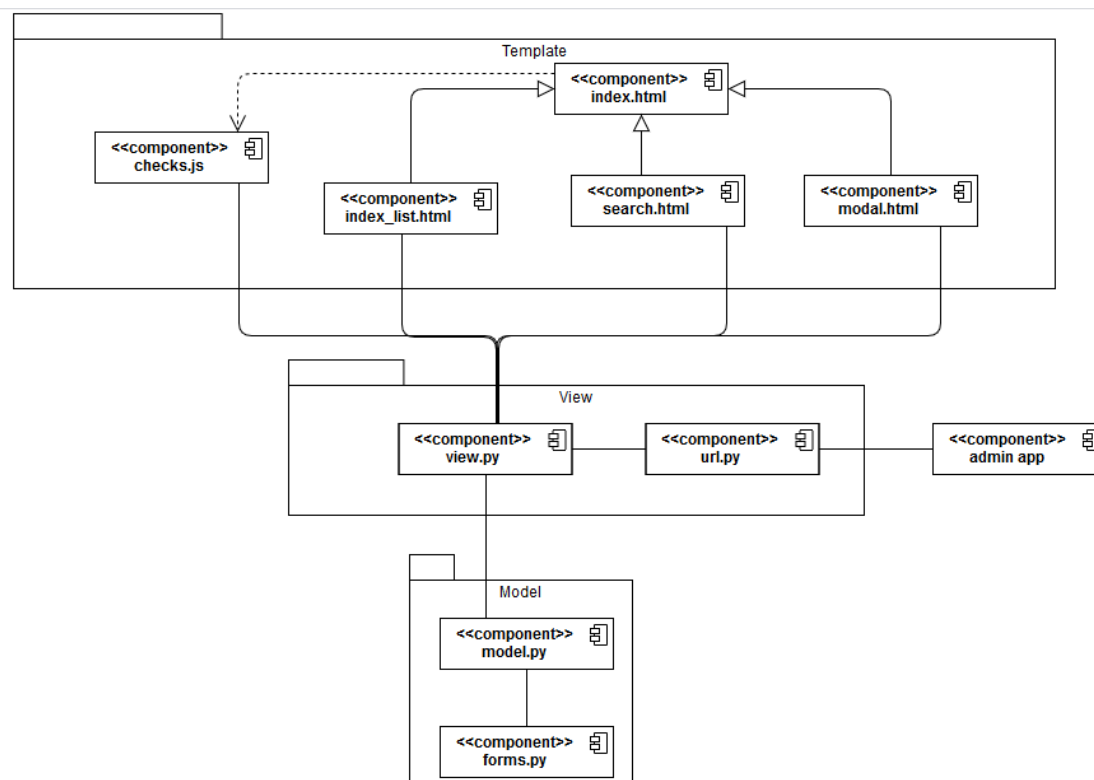


Figura 17 Diagrama de componentes da Biblioteca Digital.

Todos os dados foram armazenados numa estrutura apresentada na Figura 18, que tem os conceitos básicos do modelo Entidade-Relacionamento definidos na camada *Model*. A camada contém além disso, os formulários que permitirão inserir cada uns dos conceitos conduzidos na pesquisa persistentemente.

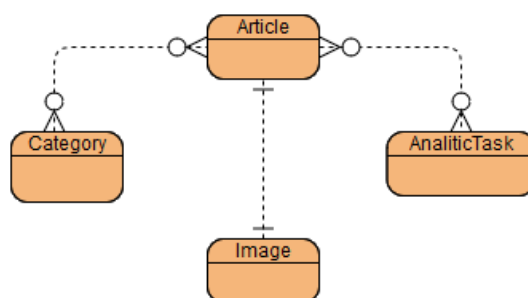


Figura 18 Modelo Entidade-Relação.

A estrutura de armazenamento da Figura 18 e as especificações na hora de guardar os dados permite-nos agilizar as buscas e estruturar melhor o conhecimento.

4.4 Uso da *MinViewREDL*

Os novos artigos que enriquecerão a base de conhecimento da *MinViewREDL* serão inseridas a partir do usuário gerenciador, quem os catalogará de acordo com os dados proporcionados pelo(s) usuário(s) contribuinte(s) mesmo como em “*Survey of Text Visualization Techniques*” (KUCHER, KERREN, 2015). A Figura 19 exhibe os dados principais para armazenar novos artigos, com especial ênfase em: título, autor(es), data de publicação, tarefa(s) analítica(s) da mineração, categoria(s) de visualização e a imagem do gráfico produzido neles.

Por outro lado, o usuário pesquisador pode acessar a interface principal (Figura 20), camada *Template*, onde pode ver todos os artigos que a biblioteca possui. Essa Figura 20 reflete no menu a direita o relacionamento (*Mining Analytic Tasks*, *Visualization Categories*) mostrado na Figura 18 onde é possível desativar ou marcar para fazer alguma filtragem tanto pelas categorias de visualização quanto pelas tarefas analíticas onde serão apresentados os artigos que cumprem com essas características.

The screenshot shows the 'Administração da Biblioteca' interface. At the top, there is a navigation bar with 'Início', 'Appdgitlib', 'Articles', and 'Adicionar artigo'. The main form is titled 'Adicionar artigo' and contains the following fields:

- Author:** A text input field.
- Title:** A text input field.
- Doi:** A text input field.
- Text:** A large text area for the article content.
- Published date:** A date selector.
- Categories:** A dropdown menu with options: DISTRIBUTION, CORRELATION, RANKING, PART OF WHOLE, FLOW, TEXT, 3D, OTHER. A plus sign is visible to the right of the list.
- Tasks:** A dropdown menu with options: LABELED, CLUSTERING, HIGHLIGHT, SUMMARIZED, SIMILARITY, OTHER. A plus sign is visible to the right of the list.
- Image:** A field for uploading an image, with a plus sign to the right.

At the bottom right of the form, there are three buttons: 'Salvar e adicionar outro(a)', 'Salvar e continuar editando', and 'SALVAR'.

Figura 19 Interface para adicionar artigos novos.


DIGITAL LIBRARY

A Visual Survey Of Text Mining Visualization For Requirement Engineering

SEARCH

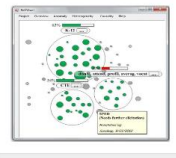
Manage visualization

EA-Miner: A Tool for Automatin...




2005

ReCVisu: A tool for clustering-b...




2009

Rig—Na environment for soft...



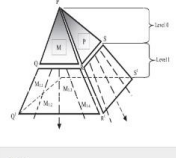
2010

Improved representation of tra...



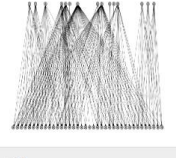
2011

Visualizing requirements: A thr...



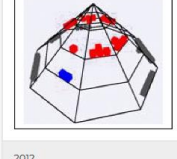
2011

Text Mining Support for Softwa...



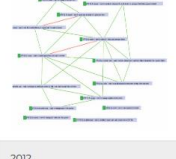
2011

Visualizing requirements in dis...



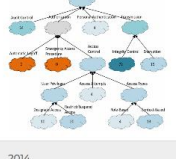
2012

Requirements traceability mat...



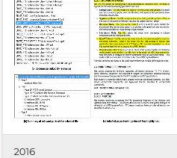
2012

Automated extraction and visu...




2014

Mining Requirements Known...




2016

Current research themes in sof...




2016

ECrits—Visualizing Support T...




2017

Let's Hear it from RETTA: A Req...




2017

Perspectives on User Story Bas...



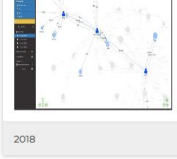
2017

ELICA: An Automated Tool for ...



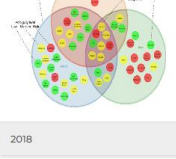
2018

The Interactive Narrator Tool: EF...



2018

Pinpointing Ambiguity and Inc...



2018

Mining Analytic Tasks

- LABELED
- CLUSTERING
- HIGHLIGHT
- SUMARAZED
- SIMILARITY
- OTHER

Visualization Categories

- CORRELATION
- RANKING
- PART OF WHOLE
- FLOW
- TEXT
- 3D
- OTHER

Copyright E.R. PUC-RIO 2019

Figura 20 Interface inicial da Biblioteca Digital.

É possível buscar por uma frase ou palavra específica que pertença ao título ou pelo nome do(s) autor(es), no campo de busca que apresenta a Figura 21, seguindo o mesmo comportamento da ferramenta base (KUCHER, KERREN, 2015).

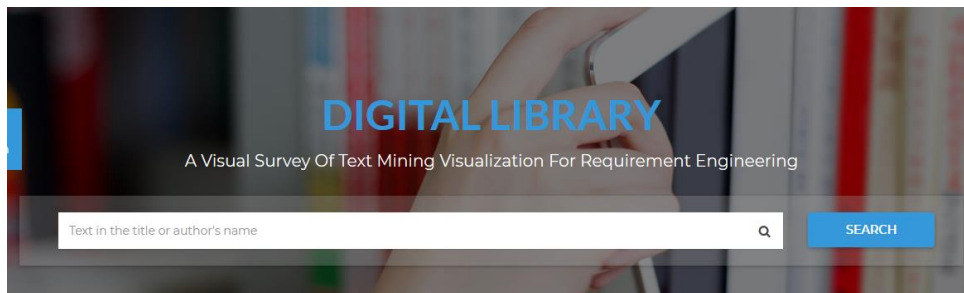


Figura 21 Buscar artigo.

Os detalhes do artigo podem ser apresentados ao selecioná-lo, como: autor(es), ano de publicação, categoria(s) de visualização e tarefa(s) analítica(s) presentes nele, e o vínculo para a leitura completa do artigo. Um exemplo é exibido na Figura 22.

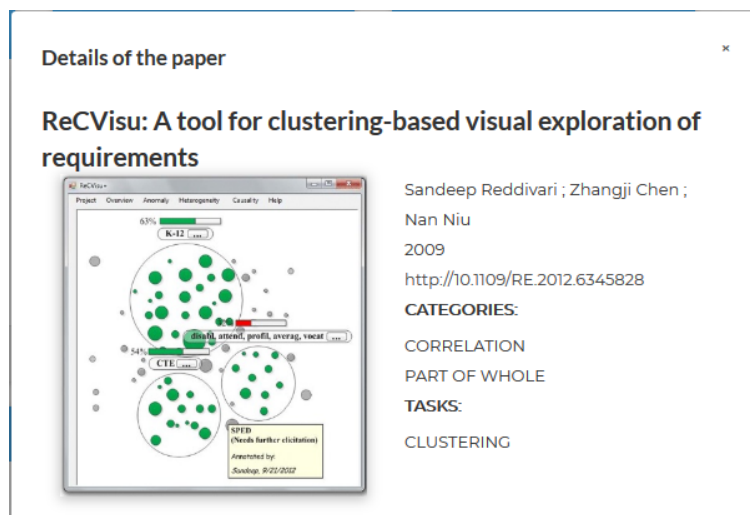


Figura 22 Detalhes do artigo.

4.5 Sumário

A biblioteca digital apresentada neste capítulo permitiu organizar o relacionamento entre as tarefas de mineração e as categorias de visualização usadas na ER de modo a possibilitar aos usuários uma busca estruturada.

5 Avaliação da estratégia

Este capítulo descreve uma avaliação da biblioteca digital, a fim de analisar a abordagem da Seção 3, que visa relacionar as categorias de visualização às tarefas de mineração que serão usadas na ER. O objetivo perseguido foi medir a relação qualidade-tempo na obtenção de conhecimento através de duas atividades diferentes. Como resultado os participantes transferiram a suas experiências respondendo um questionário.

5.1 Objetivo

Realizou-se uma avaliação retomando a pergunta e os objetivos definidos nesta pesquisa: *Como melhorar o processo de obtenção de conhecimento sobre visualização de texto minerado na Engenharia de Requisitos?*

De acordo com a estratégia desenvolvida, a seleção de visualizações é possível na MinViewREDL, de acordo com a Tabela 2 que a figura 18 implementa, ou seja, dada uma categoria de visualização ou tarefa analítica (Figura 20) permite recuperar artigos que atendam esse relacionamento.

Para conhecer se a estratégia ajuda nessa seleção foi projetado uma avaliação com o seguinte **objetivo**: Medir a relação qualidade-tempo, no processo de obtenção de conhecimento sobre a relação de mineração de texto com a visualização empregada, usando a estratégia proposta.

Para cumprir este objetivo é preciso entender que o processo de obtenção de conhecimento é evidenciado sobre o estado da arte na descoberta das visualizações de texto minerado que são usadas para extrair requisitos. Em relação a isso, é desejável para o engenheiro de requisitos saber se o uso da biblioteca é eficaz.

A avaliação foi feita medindo a quantidade de artigos que cada participante consegue encontrar em um tempo fixo com duas formas diferentes: uma busca livre e outra usando a *MinViewREDL*.

Foram definidos os passos principais a executar pelo participante:

Passo 1: Procurar em 10 minutos pesquisas que incluam visualizações de texto minerado para Engenharia de Requisitos. Pode usar o método com que sinta comodidade.

Passo 2: Procurar em 10 minutos pesquisas que incluam visualizações de texto minerado para Engenharia de Requisitos usando a biblioteca digital *MinViewREDL*.

Finalmente o participante respondeu um questionário onde evidenciou as suas considerações. O questionário preenchido após de execução da avaliação (passos 1 e 2 descritos) é evidenciado na Tabela 3. As perguntas visam a avaliar, de modo geral, a eficácia da Biblioteca desenvolvida.

Tabela 3 Questionário de avaliação.

NO.	PERGUNTA	RESPOSTA
P1.	De modo geral, eu achei a ferramenta...	Muito ruim _ (1) _ (2) _ (3) _ (4) _ (5) Muito boa
P2.	Achei que o uso da ferramenta foi...	Muito difícil _ (1) _ (2) _ (3) _ (4) _ (5) Muito fácil
P3.	Achei que o objetivo que persegue a ferramenta foi ...	Muito ruim _ (1) _ (2) _ (3) _ (4) _ (5) Muito bom
P4.	Entender as classificações propostas pela ferramenta foi ...	Muito difícil _ (1) _ (2) _ (3) _ (4) _ (5) Muito fácil
P5.	Você encontrou o que procurava?	<input type="checkbox"/> Não <input type="checkbox"/> Sim <input type="checkbox"/> Mais ou menos
P6.	Você achou a navegação fácil e intuitiva?	<input type="checkbox"/> Não <input type="checkbox"/> Sim <input type="checkbox"/> Outra
P7.	Indique fatores positivos que encontrou na ferramenta.	
P8.	Indique fatores negativos que encontrou na ferramenta.	
P9.	Pode dar alguma sugestão de melhoria da ferramenta?	

O questionário aborda, um tanto, aspectos sobre Conhecimento adequado, Capacidade de ser usado (Operabilidade) e Estética da interface do usuário (“ISO 25010”, 2019). O que é traduzido a:

- Conhecimento adequado: capacidade do produto que permite ao usuário entender se o software é adequado para suas necessidades. Este aspecto está refletida diretamente nas perguntas P3, P4 e P5 do questionário de avaliação aplicado.
- Capacidade de utilização (Operabilidade): capacidade do produto que permite ao usuário operá-lo e controlá-lo com facilidade. Principalmente evidenciada nas perguntas P1 e P2 do questionário de avaliação.
- Estética da interface do usuário: capacidade da interface do usuário para agradar e satisfazer a interação com o usuário. Abordada pela pergunta P6.

Foram elaboradas outras perguntas para que os participantes expressarem as suas opiniões e contribuir no processo de melhoria continua da abordagem como a P7, P8 e P9 da Tabela 3.

Com os passos descritos pretendeu-se cobrir os objetivos traçados para avaliar a estratégia, iniciando com a rapidez para encontrar sugestão de visualização para texto minerado.

5.2 Seleção dos participantes

A amostra para nossa avaliação foi determinada pela quantidade de convites aceitos por engenheiros de software com conhecimentos da ER, em um total de 17, nossa população de avaliadores. Nessa seleção foi solicitado aos participantes informar a quantidade de anos de experiência atuando com mineração de texto para serem ordenados ascendentemente de acordo com este fator.

Para a seleção dos participantes na Avaliação Piloto foi aplicado um método de Exemplificação sistemático já que permite cobrir todos os tipos de unidade, neste caso, poucos, médio ou muitos anos de experiência na mineração de texto. Dessa maneira da lista de 17 participantes se quer uma mostra de 3 sujeitos (n). Para isso ordenamos os elementos da população (N) começando por 0 anos de experiência

em forma crescente e obtemos o coeficiente de elevação k aplicando a equação 1 (SEOANE et al., 2007):

$$k = N / n \quad (1)$$

A origem da seleção começou pelo primeiro indivíduo (selecionado aleatoriamente) da lista ordenada de participantes e sucessivamente foi tomado pelo indivíduo que está a k lugares da última escolha. Permitiu-se contar com a participação de 3 indivíduos de níveis diferentes de conhecimento em mineração de texto que conformaram o grupo para a Avaliação piloto.

Para a Avaliação Final contou-se com os restantes 14 participantes. O grau de escolaridade deles é composto por 92,85% de pós-graduandos. O restante tem grau Superior, mas amplo conhecimento trabalhando na área de mineração de texto. A profissão de 50% deles é composta por Estudantes, 24,29% são Pesquisadores e os restantes Desenvolvedores de Tecnologia de Informação, experientes na área de ER.



Figura 23 Caracterização pelos anos de experiencia.

Cabe destacar a partir da caracterização dos participantes que alguns possuem menos de 1 ano de experiencia na área de mineração de texto e outros contam somente conhecimentos teóricos, quem foram sujeitos de passos adicionais que serão explicados na seção da execução.

5.3 Execução

A execução foi feita em dois subprocessos, como exhibe a Figura 24, o primeiro foi definido como Avaliação Piloto e o segundo como Avaliação Final. A

autora deste trabalho e responsável desta pesquisa participou como moderadora nestes subprocessos com o objetivo de explicar as atividades a serem realizadas e esclarecer dúvidas, mas sem influenciar nas ações executadas.

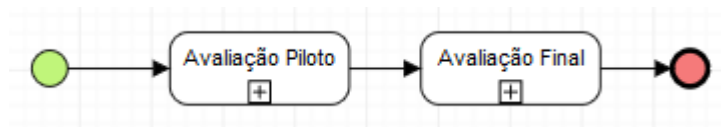


Figura 24 Processo de avaliação.

A Avaliação Piloto tinha a missão de identificar lacunas ou ambiguidades no processo que seria realizado como Avaliação Final. A partir disso foi definida para ganhar em conhecimento na execução do instrumento de avaliação. Dessa forma seguiu o processo descrito na Figura 25 para cada participante individualmente e contribuiu a melhorar a compreensão da dinâmica aplicada.

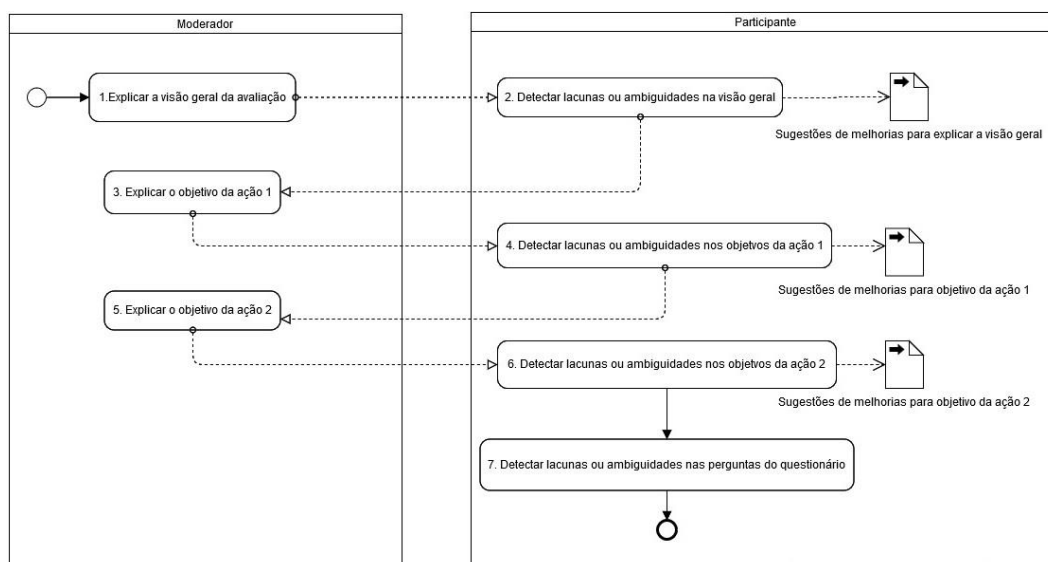


Figura 25 Subprocesso de Avaliação Piloto.

Na atividade 1 da Figura 25: foi explicada a motivação da avaliação assim como os objetivos gerais a atingir. O participante da Avaliação Piloto, como atividade consequente (atividade 2), deu a sua opinião do que deve ser modificado nessa visão para que seja claramente entendido. Foi necessário adotar medidas para quem não estava familiarizado com a mineração de texto. Para isso foi criado um documento de Conteúdos de Ajuda para responder questionário (Veja Apêndice III) foi disponibilizado e assim, o nível do conhecimento não constituiu uma ameaça na hora de enfrentar as ações do Avaliação Final.

Na atividade 3 o participante foi informado da ação que devia ser executada, neste caso o passo 1. Essa atividade incluía procurar artigos usando um método de livre escolha em um tempo de 10 minutos. Para a detecção de inconsistências o participante executou a ação e como saída da atividade 4, propôs sugestões de tempo apropriado para execução da atividade.

A atividade 5 atingiu ao passo 2, para comparar a melhora na seleção de visualizações, neste caso a partir da busca feita na *MinViewREDL*, que como foi dito, provê visualizações para mineração de texto na ER. Com a execução da atividade 6, o participante fez anotações de acordo com o entendimento da busca na biblioteca digital. A partir dessa atividade foi sugerido, como resultado, incrementar a Avaliação Final com mais uma atividade. A nova atividade fundamenta-se na prévia consulta na ferramenta base “*Survey of Text Visualization Techniques*” (KUCHER, KERREN, 2015). Mesmo que ela não incluía pesquisas para a ER, o participante pode conseguir comparar o funcionamento de ambas ferramentas.

Por último, com a atividade 7, o participante respondeu às perguntas que mediram eficácia da *MinViewREDL*. Com esta atividade conseguiu-se resumir o processo executado.

O subprocesso de Avaliação Final foi executado por outro grupo de participantes. Cada um deles individualmente dedicou 40 minutos em meia com a modeladora para efetuar as atividades da avaliação. Aconteceu na sala 525 do Departamento de Informática de PUC-Rio, sem interrupções que atentassem contra o desempenho da avaliação. Foi destinado um computador para o participante e a presença do moderador nas atividades foi para explicar as ações, esclarecer dúvidas e fazer anotações do processo.

As atividades realizadas são detalhadas na Figura 26, definidas de acordo com os passos descritos anteriormente. As mesmas foram executadas imaginando o cenário de ser engenheiro querendo minerar texto para extrair requisitos e precisando conhecer o estado da prática, principalmente os artigos que desenvolveram visualizações com esse fim.

Foi a explicação da visão geral que deu começo a Avaliação Final, mesmo como na Avaliação Piloto. Seguidamente foi proporcionado um conteúdo de ajuda para aqueles participantes com baixos conhecimentos na área de mineração de texto, refletido na atividade 1.1 da Figura 26.

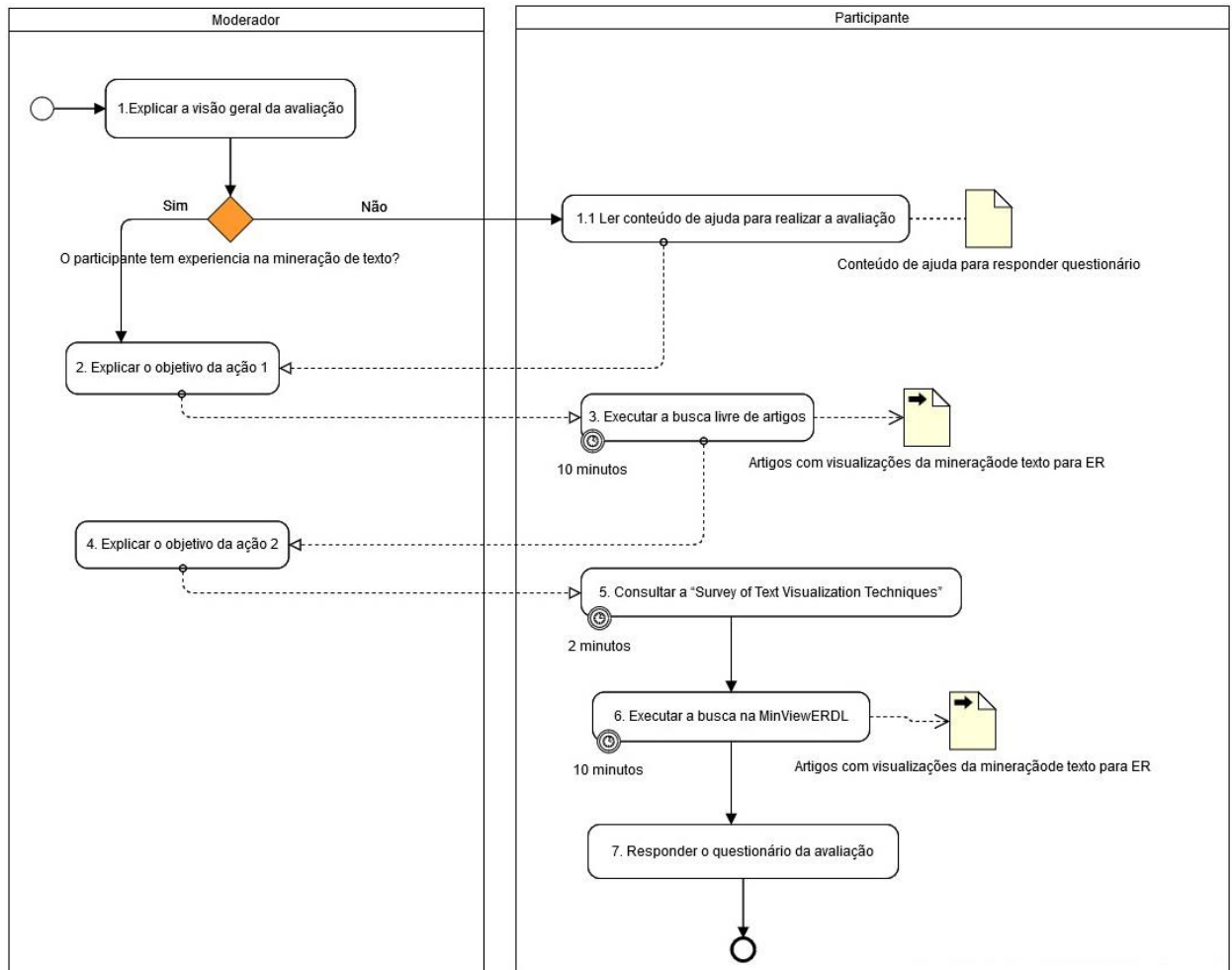


Figura 26 Subprocesso do Avaliação Final.

A atividade 2 correspondeu a explicação do resultado esperado pela atividade de busca no passo 1 descrito com anterioridade com limite de tempo de 10 minutos. Solicitou-se a busca de artigos que incluíssem visualizações de texto minerado para a ER usando um método livre. A atividade 3 é o cumprimento do participante da atividade explicada.

Na atividade 4, após a explicação da próxima ação, corresponde a execução da busca na *MinViewERDL* (biblioteca digital descrita na Seção 4), mas antes disso foi orientado consultar a "*Survey of Text Visualization Techniques*".

A consulta da atividade 5 foi com tempo curto de 2 minutos acessando a “<http://textvis.lnu.se/>”. É justificada esta atividade principalmente com o objetivo da familiarização com o seu funcionamento, já que a *MinViewREDL* foi desenvolvida seguindo essa mesma ideia. Nesse momento foi explicada a principal diferença em relação as duas ferramentas, sabendo que ambas estão orientadas na organização do conhecimento das visualizações de texto minerado, mas a *MinViewREDL* é especializada na ER. Esta atividade (atividade 5) foi produzida após da Avaliação Piloto considerando-se como útil para melhorar o entendimento do processo.

Com um conhecimento base obtido da atividade 5 a atividade 6 refere-se ao passo 2 da avaliação, a busca na biblioteca desenvolvida. A partir da experiência na atividade 6, a atividade 7 recolheu a valoração da avaliação.

5.4 Análise dos resultados

Como foi explicado anteriormente, os objetivos foram materializados em dois passos principais e divididos em atividades executada pelos participantes. Foi medido pela quantidade de resultados encontrados na busca livre, que ajudaram com a identificação da mineração de texto para ER e que ao mesmo tempo, exibiram suas visualizações.

Foram feitas anotações pela moderadora, a partir da observação do processo executado por cada participante. De acordo com isso pode-se ressaltar que os participantes usaram palavras chaves tanto em inglês quanto em português e a máquina de busca de *Google Acadêmico* como ferramenta de amplo espectro y acceso livre. Dois participantes com o decorrer do tempo, mudaram de buscador para *IEEE Xplorer Digital Library*. Independentemente das palavras chaves que foram usadas, entre elas: *software visualization, requirement, NLP, text mining, elicitation* os resultados dessa busca livre foram limitados. Sem deixar de mencionar que muitos participantes (9) não conseguiram achar nenhum dentro do limite de tempo regulamentado. Corroboraram a hipóteses das ainda insuficientes pesquisas desenvolvidas para a ER que incluem mineração de texto e apresentam

as visualizações desenvolvidas além da força tarefa para pesquisadores de vincular três áreas em uma busca (ER, Mineração de texto e Visualização).

Do total de trabalhos encontrados: foram divididos em teses, artigos sem visualização (resumos), revisão sistemática de literatura, artigos só de mineração com visualização, ou mineração para outras tarefas de desenvolvimento de software.

Um total de 8 artigos foram encontrados na busca livre pelos participantes, 7 não tinham sido captados nos 14 artigos da nossa busca bibliográfica. Deles três (3) foram inseridos como novas contribuições e outros quatro (4) estão em processo de estudo.

A média dos artigos encontrados por participantes na busca livre foi de 0,64 artigos de interesse enquanto a busca na *MinViewREDL* foi de 5,33 por participantes. Indica que o uso da biblioteca permitiu encontrar uma quantidade maior de resultados já que, armazena somente visualizações de texto minerado para ER.

Obteve-se, também, a opinião dos participantes na busca livre. A maioria expressou surpresa com a dificuldade do trabalho multidisciplinar e a novidade que representa a mineração e a visualização para a ER. Foi muito importante para eles conhecer as tarefas de mineração usadas na ER, algo um tanto oculto pela literatura atualmente.

Com relação às considerações dos participantes sobre a eficácia do processo de aquisição de conhecimento, foi realizada uma análise com base nas respostas do questionário proposto para a avaliação e descrito com base nos aspectos gerais projetados.

Conhecimento adequado

A parte do questionário que aborda se o software está adequado as necessidades do usuário, avalia o cumprimento do objetivo para o qual o software foi criado. Dessa forma, as perguntas foram as P3, P4 e P5. Cabe ressaltar o seguinte:

P3. Do total de participantes, 13 (92,86 %) classificaram como "Muito Bom" o objetivo definido para a ferramenta *MinViewREDL*. Apenas 1 participante classificou esse objetivo como "Bom".



Figura 27 Estatísticas para a pergunta 3.

P4. Entender as classificações de tarefas de mineração e categorias de visualização proporcionadas para um participante (1) nem foi fácil nem difícil, avaliação media, mais um (1) considerou que foi “Fácil” e 85,71% entenderem que foi “Muito Fácil”.

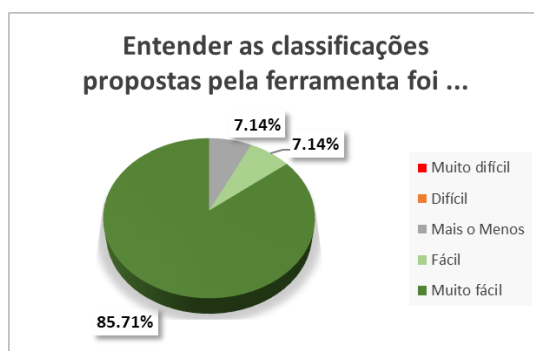


Figura 28 Estatísticas para a pergunta 4.

P5. Alguns dos participantes procuravam encontrar os mesmos artigos da busca livre considerando que não relacionaram as três áreas. O 64,29% respondeu que “Sim” encontrou o que procurava, o resto manifestou neutralidade (“Mais o menos”).

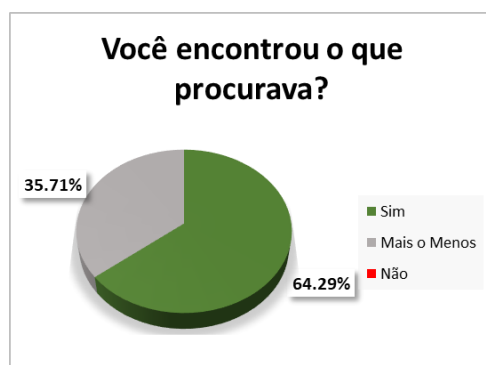


Figura 29 Estatísticas para a pergunta 5.

Cabe destacar que o resultado esteve influenciado por o uso de expressões complexas no campo de busca com conectores AND e OR que não foi concebido para esta versão.

Acerca do nível de entendimento das funcionalidades da biblioteca pelo usuário tem um índice de resultado alto. As maiores porcentagens estão localizadas em avaliações positivas e nenhum registro de avaliação negativa como é apresentado na Tabela 4.

Tabela 4 Resumo de resultados dos aspetos sobre o conhecimento adequado.

<i>Conhecimento adequado</i>				
	P3	P4		P5
1				
2				
3		7,14 %	<i>Não</i>	
4	7,14 %	7,14 %	<i>Mais o menos</i>	35,71 %
5	92,86 %	85,71 %	<i>Sim</i>	64,29 %

Capacidade de utilização (Operabilidade)

A operabilidade da *MinViewREDL* é medida destacando principalmente o grau de facilidade com a qual o usuário a opera. Foram incluídas as perguntas P1 e P2 sob esta análise, ressaltando o seguinte:

P1. De modo geral o 78,57% dos interrogados achou que é uma ferramenta “Muito Boa” e a outra parte considerou que é “Boa”.

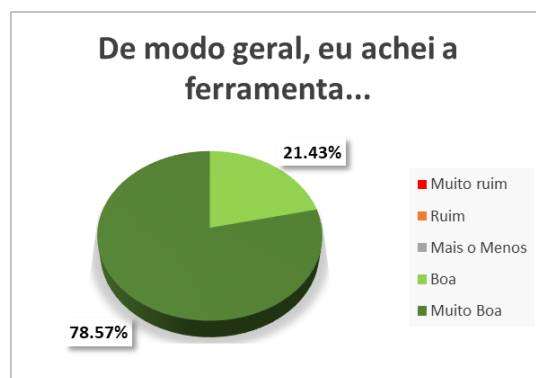


Figura 30 Estatísticas para a pergunta 1.

P2. Com relação à opinião do uso um 57,14% respondeu que é “Muito fácil” e o resto concordou que é “Fácil”.

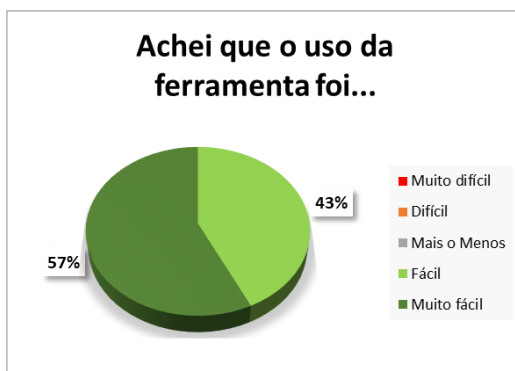


Figura 31 Estatísticas para a pergunta 2.

Um fator de peso para esta pergunta é o nível de conhecimento dos participantes com os termos empregados, já que são muito específicos da área de visualização e mineração, com os quais não precisam estar familiarizados. Mesmo assim, foi transmitido para o usuário a utilidade da *MinViewREDL*.

Evidenciou-se na Tabela 5 que os resultados da questão 1 estão agrupados nas avaliações de Bom e Muito Bom. Por outro lado, na pergunta 2, foi dito que, em geral, o uso da *MinViewREDL* está entre muito fácil e fácil. Destaca-se que não foram registradas avaliações negativas para a operabilidade da biblioteca.

Tabela 5 Resumo de resultados da Operabilidade.

	<i>Capacidade de utilização</i>	
	P1	P2
1		
2		
3		
4	21,43 %	42,86 %
5	78,57 %	57,14 %

De acordo com a escala estabelecida na elaboração do questionário, cinco correspondem a uma avaliação positiva maior e vai em diminuição.

Estética da interface do usuário

Compreende à capacidade para a interação satisfatória e mede o grau de facilidade para o uso, evidenciando-se nas respostas da pergunta P6.

P6. A biblioteca digital é Muito intuitiva e fácil de usar para o 92.86% do total de participantes. Comprovado pela quantidade de comentários positivos que o refletem e destacam características da sua interface e dos termos que ela manipula.

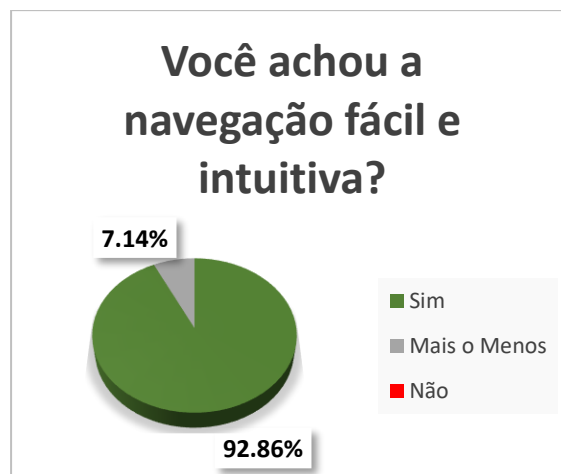


Figura 32 Estatísticas para a pergunta 6.

Foram recolhidos comentários para fatores positivos, negativos e sugestões com a finalidade de seguir melhorando a estratégia e a *MinViewREDL*. A seguir são expostos a modo de resumos.

P7. Indique fatores positivos que encontrou na ferramenta

Os registros no questionário evidenciam alusões a características de elegância estética, cores harmoniosas, tamanho legível de texto e botões, auto explicável, amigável como alguns dos fatores recorrentes. Outro fator positivo a destacar é a facilidade para realizar as buscas, uma ferramenta simples e objetiva. Agrupa pesquisas focadas para engenharia de requisitos que destaca o valor para os objetivos traçados neste trabalho.

P8. Indique fatores negativos que encontrou na ferramenta.

Entre os fatores negativos faltou explicação de como fazer as buscas já que o mecanismo não é tão avançado (por exemplo, não permite uso de conectivos AND e OR para facilidade do pesquisador). A busca atual é só pelo texto no título e autores (talvez poderia existir um filtro para o resumo dos artigos). Conta com poucos artigos e não era perceptível o vínculo para acessar cada um deles presente na janela dos detalhes do artigo. Uso deficiente de *tooltip* para ajudar na explicação sobre as tarefas de mineração e as categorias de visualização que permitirá melhor compreensão para usuários novos no trabalho com a mineração na ER e a visualização.

Finalmente a pergunta P9 expõe sobre as sugestões de melhoria: *P9. Pode dar alguma sugestão de melhoria da ferramenta?*

- Avaliar a possibilidade de um corretor de texto no campo de busca.
- Colocar de alguma maneira como o usuário poderia fazer as pesquisas.
- Melhorar mecanismo de busca para torna-lo mais complexo permitindo o uso de conectores.
- Talvez apresentar categorias de temas ou tarefas dos artigos ao abrir a ferramenta antes de apresentar todos os artigos.
- Adicionar o resumo de cada artigo e *Keywords* nas janelas dos detalhes dos artigos.
- Adicionar uma categoria baseada nas atividades de requisitos.
- Criar um menu de ajuda com informações uteis assim como um espaço para o próprio autor envie contribuições como acontece no “*Survey of Text Visualization Techniques*”.
- Detalhar melhor o nível das tarefas de NLP de cada trabalho, por exemplo, *Labeled* a nível de sentença.

Cabe destacar que as sugestões dos participantes permitirão melhorar o desenvolvimento da biblioteca digital no sentido de dar comodidade na pesquisa e no seu uso.

Os artigos produzidos das buscas na avaliação serão parte do enriquecimento da base de conhecimento. Destacou-se uma sugestão interessante que pode ser estudada, como é o caso de “melhorar o nível das tarefas de NLP de cada trabalho”. Refere-se a mostrar para o usuário aqueles tipos de dados categóricos que são visualizados.

Pode-se enfatizar que os comentários produzidos pelos participantes indicam que a solução apresentada nesta dissertação é coerente com o objetivo desejado.

5.5 Sumário

Com o desenho e execução da avaliação foi possível medir a relação qualidade-tempo na obtenção de conhecimento a partir do uso da “*MinViewREDL*”.

A abordagem resultante das análises das categorias de visualização e as tarefas analíticas de mineração traduzidos na biblioteca digital é facilmente compreendido a partir da avaliação feita pelos usuários, com dados estatístico positivos. As estatísticas também mostraram o grau de satisfação em relação à facilidade com a qual o usuário opera a biblioteca digital, mesmo que lhe falte maturidade.

6 Conclusão

6.1 Resumo

Com a pesquisa realizada, evidenciou-se o aumento do uso da mineração de textos e as diferentes formas de representação de seus resultados através da consulta na ferramenta “*Survey of Text Visualization Techniques*” (KUCHER, KERREN, 2015). Apresentou-se que a incidência de mineração na ER também está aumentando. No entanto esse aumento ainda é insuficiente devido ao crescimento dos dados que precisam ser manipulados em tarefas de elicitação. Nesta dissertação elicítamos conhecimento relativo ao uso das tarefas de mineração e técnicas de visualização utilizadas na ER.

Retomando a pergunta de pesquisa: *Como melhorar o processo de obtenção de conhecimento sobre visualização de texto minerado na Engenharia de Requisitos?* Pode-se concluir que:

A partir da consulta da ferramenta “*From Data to Viz*”(HEALY, [s.d.]) foram identificadas as categorias de visualização para as tarefas de mineração de texto na ER. Com o relacionamento da mineração e visualização na ER, a partir dos artigos encontrados na busca de bibliografia o engenheiro obtém informações para se enfrentar os desafios de usar mineração de texto para extrair requisitos.

A sugestão das visualizações foi abordada nesta pesquisa a partir da análise de dados categóricos proporcionados pela mineração e a correlação estabelecida entre categoria de visualização e tarefa de mineração. Os relacionamentos foram evidenciados pelos resultados da busca de bibliografia e que constituem uma base de conhecimento.

Para garantir a persistência deste relacionamento proposto, foi desenvolvido o software “*MinViewREDL*”, fundamentada em “*Survey of Text Visualization Techniques*” (KUCHER, KERREN, 2015), que condensou as classificações, as relações e exibe as pesquisas que materializam como é possível usar mineração de

texto na Engenharia de Requisitos e quais visualizações podem ajudar a exibir os resultados.

Com a avaliação através da medição da qualidade-tempo na obtenção de conhecimento a partir do uso da “*MinViewREDL*”, identificou-se que o processo de busca baseado na pré-classificação de tarefas de mineração e categorias de visualização traz agilidade. Isso é possível porque a ferramenta melhora a organização do conhecimento. Isso foi evidenciado nas considerações dos participantes da avaliação.

6.2 Contribuições

Como contribuição principal pode-se destacar a relação existente e descrita neste trabalho entre as categorias de visualização e tarefas de mineração para a ER, delimitando as que são usadas especificamente nesta área e que foram contempladas a partir do refinamento da busca bibliográfica.

A ajuda ao processo de elicitação é evidenciado a partir do desenvolvimento da *MinViewREDL*, que organiza os artigos que usam mineração de texto com categorias de visualizações para ER. Com o armazenamento e estruturação do conhecimento evidenciado na biblioteca digital proporciona-se a primeira aproximação de uma base de conhecimento que inclui a referência onde foram desenvolvidas as visualizações e a tarefa de mineração executada para uma área específica: ER.

A infraestrutura utilizada para geração da *MinViewREDL* está disponibilizada em: https://github.com/AdilaMB/PPF_2019.git. Dessa forma, nosso trabalho pode evoluir tanto com acréscimo de nova literatura, mas também através de novas funcionalidades em função do uso de código aberto.

6.3 Trabalhos futuros

Nesta pesquisa foi criada a infraestrutura para coletar as investigações que relacionam a visualização às tarefas de mineração que são realizadas na ER. No futuro, sugerimos os seguintes aperfeiçoamentos:

- Detalhar o nível das tarefas de mineração de acordo com os dados categóricos de cada artigo.
- Evoluir a biblioteca digital incorporando-lhe as categorias focadas nas atividades de requisitos.
- Incluir seção de ajuda com informações úteis.
- Incorporar o cadastro de novas sugestões de artigos através da própria biblioteca.

A arquitetura desenvolvida oferece a possibilidade da inclusão de componentes fornecido por terceiros. Por exemplo, a inclusão de componentes para gerar visualizações com base no fornecimento de texto minerado.

Vislumbra-se também a possibilidade de generalizar a arquitetura como *framework* para associar novos conceitos do domínio da ER na criação de bibliotecas específica.

7 Referências bibliográficas

ABAD, Z. S. H.; NOAEEN, M.; RUHE, G. **Requirements Engineering Visualization: A Systematic Literature Review**. 2016 IEEE 24th International Requirements Engineering Conference (RE). **Anais...** In: 2016 IEEE 24TH INTERNATIONAL REQUIREMENTS ENGINEERING CONFERENCE (RE). set. 2016

CALVO, J. M. G. Consideraciones sobre la palabra como unidad lingüística. **Revista española de lingüística**, v. 12, n. 2, p. 375–410, 1982.

DIEHL, S. **Software Visualization: International Seminar Dagstuhl Castle** Springer Science & Business Media, , May 20-25.

DIEHL, S. **Software Visualization: International Seminar Dagstuhl Castle, Germany, May 20-25, 2001 Revised Papers**. Germany: Springer Science & Business Media, 2002. v. 2269

DJANGO SOFTWARE FOUNDATION. **Django**. Lawrence, Kansas: [s.n.].

DUARTE, D. et al. **Collaborative Requirements Elicitation with Visualization Techniques**. 2012 IEEE 21st International Workshop on Enabling Technologies: Infrastructure for Collaborative Enterprises. **Anais...** In: 2012 IEEE 21ST INTERNATIONAL WORKSHOP ON ENABLING TECHNOLOGIES: INFRASTRUCTURE FOR COLLABORATIVE ENTERPRISES. jun. 2012

FEATHER, M. S. et al. **Experiences using Visualization Techniques to Present Requirements, Risks to Them, and Options for Risk Mitigation**. 2006 First International Workshop on Requirements Engineering Visualization (REV'06 - RE'06 Workshop). **Anais...** In: 2006 FIRST INTERNATIONAL WORKSHOP ON REQUIREMENTS ENGINEERING VISUALIZATION (REV'06 - RE'06 WORKSHOP). set. 2006

FENG, L.; CHIAM, Y. K.; LO, S. K. **Text-Mining Techniques and Tools for Systematic Literature Reviews: A Systematic Literature Review**. 2017 24th Asia-Pacific Software Engineering Conference (APSEC). **Anais...** In: 2017 24TH ASIA-PACIFIC SOFTWARE ENGINEERING CONFERENCE (APSEC). dez. 2017

GOTEL, O. C. Z.; MARCHESE, F. T.; MORRIS, S. J. **On Requirements Visualization**. Second International Workshop on Requirements Engineering Visualization (REV 2007). **Anais...** In: SECOND INTERNATIONAL WORKSHOP ON REQUIREMENTS ENGINEERING VISUALIZATION (REV 2007). out. 2007

HEALY, Y. H. AND C. **From data to Viz | Find the graphic you need**. Disponível em: <data-to-viz.com>. Acesso em: 21 maio. 2019.

ISO 25010. Disponível em: <<https://iso25000.com/index.php/en/iso-25000-standards/iso-25010>>. Acesso em: 5 fev. 2020.

KHAN, T. et al. **Visual analytics of software structure and metrics**. 2015 IEEE 3rd Working Conference on Software Visualization (VISSOFT). **Anais...** In: 2015 IEEE 3RD WORKING CONFERENCE ON SOFTWARE VISUALIZATION (VISSOFT). set. 2015

KUCHER, K.; KERREN, A. **Text visualization techniques: Taxonomy, visual survey, and community insights**. 2015 IEEE Pacific Visualization Symposium (PacificVis). **Anais...** In: 2015 IEEE PACIFIC VISUALIZATION SYMPOSIUM (PACIFICVIS). abr. 2015

LATHAM, A. **orthogonal/Lyrics-Commander**. [s.l: s.n.].

LEITE, J. C. S. P. **Livro Vivo : Engenharia de Requisitos: A Necessidade da Elicitação**, 2007. Disponível em: <<http://livrodeengenhariaderequisitos.blogspot.com/2007/09/necessidade-da-elicitaao.html>>. Acesso em: 19 maio. 2019

Mac Morpho. Disponível em: <<https://kaggle.com/nltkdata/mac-morpho>>. Acesso em: 4 fev. 2020.

MARTIN FOWLER et al. **Manifesto para Desenvolvimento Ágil de Software**. Disponível em: <<https://agilemanifesto.org/iso/ptbr/manifesto.html>>. Acesso em: 26 dez. 2019.

Parágrafo. In: **DICIO, Dicionário Online de Português. Porto: 7Graus**, 2019a. Disponível em: <<https://www.dicio.com.br/paragrafo/>>. Acesso em: 20 fev. 2020

PORTUGAL, R. L. Q. **MINERAÇÃO DE INFORMAÇÃO EM LINGUAGEM NATURAL PARA APOIAR A ELICITAÇÃO DE REQUISITOS**. MESTRE EM INFORMÁTICA—Rio de Janeiro, Brazil: PONTIFÍCIA UNIVERSIDADE CATÓLICA DO RIO DE JANEIRO, 19 abr. 2016.

REDDIVARI, S. et al. Visual requirements analytics: a framework and case study. **Requirements Engineering**, v. 19, p. 257–279, 1 set. 2014.

REDDIVARI, S.; CHEN, Z.; NIU, N. **ReCVisu: A tool for clustering-based visual exploration of requirements**. 2012 20th IEEE International Requirements Engineering Conference (RE). **Anais...** In: 2012 20TH IEEE INTERNATIONAL REQUIREMENTS ENGINEERING CONFERENCE (RE). set. 2012

SANTOS, R. E. et al. Técnicas De Processamento De Linguagem Natural Aplicadas Ao Processo De Mineração De Textos: Resultados Preliminares De Um Mapeamento Sistemático. **Revista de Sistemas e Computação-RSC**, v. 4, n. 2, 2015.

SCHWARTZ, C. Digital libraries: an overview. **The Journal of Academic Librarianship**, v. 26, n. 6, p. 385–393, 1 nov. 2000.

Sentença. In: **DICIO, Dicionário Online de Português. Porto: 7Graus**, 2019b. Disponível em: <<https://www.dicio.com.br/sentenca/>>. Acesso em: 20 fev. 2020

SEOANE, T. et al. Capítulo 5: Selección de la muestra: técnicas de muestreo y tamaño muestral. **SEMERGEN - Medicina de Familia**, v. 33, n. 7, p. 356–361, 1 ago. 2007.

SHIVAPRASAD, K.; REDDY, T. H. **Text mining: An improvised feature based model approach**. 2016 2nd International Conference on Applied and Theoretical Computing and Communication Technology (iCATccT). **Anais...** In: 2016 2ND INTERNATIONAL CONFERENCE ON APPLIED AND THEORETICAL COMPUTING AND COMMUNICATION TECHNOLOGY (ICATCCT). jul. 2016

SHNEIDERMAN, B. **Challenges for information visualization research: Visual quality and data quantity**. 2012 IEEE Pacific Visualization Symposium. **Anais...** In: 2012 IEEE PACIFIC VISUALIZATION SYMPOSIUM. fev. 2012

TAN, A.-H. **Text mining: The state of the art and the challenges**. Proceedings of the PAKDD 1999 Workshop on Knowledge Discovery from Advanced Databases. **Anais...**sn, 1999

WNUK, K. et al. Supporting Scope Tracking and Visualization for Very Large-Scale Requirements Engineering-Utilizing FSC+, Decision Patterns, and Atomic Decision Visualizations. **IEEE Transactions on Software Engineering**, v. 42, n. 1, p. 47–74, jan. 2016.

ZAYAS, Y. M. DEL C.; MEDEROS, A. A. L. La minería de texto: perspectiva metodológica para la realización de resúmenes documentales. **ACIMED**, v. 15, n. 5, maio 2007.

Apêndice

Apêndice I. Artigos da primeira etapa de revisão de literatura

TITULO	ATOR(ES)	TAREFA(S)	ANALÍTICA(S)	CATEGORIA(S)	ANO
				DE	
				VISUALIZAÇÃO	

ID1	Paralinguistic Recommendations for Affective Word Clouds (https://doi.org/10.1145/3301275.3302327)	Tugba Kulahcioglu Gerard de Melo	Text Summarization/Topic Analysis/ Entity Extraction Sentiment Analysis Relation/Connection Analysis	Clustering/Classification/Categorization Overview	2019
ID2	StoryPrint: An Interactive Visualization of Stories (https://doi.org/10.1145/3301275.3302302)	Katie Watson, Samuel S. Sohn, Sasha Schriber, Markus Gross, Carlos Manuel Muniz Mubbasir Kapadia	Text Summarization/Topic Analysis/ Entity Extraction Discourse Analysis Sentiment Analysis Event Analysis Trend Analysis/Pattern Analysis	Clustering/Classification/Categorization Comparison Overview Navigation/Exploration	2019

ID3	Who is the Hero, the Villain, and the Victim?: Detection of Roles in News Articles using Natural Language Techniques. (https://doi.org/10.1145/3172944.3172993)	Diego Gomez-Zara, Miriam Boon Larry Birnbaum	Text	Summarization/Topic Analysis/ Entity Extraction Sentiment Analysis	Clustering/Classification/Categorization Comparison Overview	2018
ID4	Searching for Diverse Perspectives in News Articles: Using an LSTM Network to Classify Sentiment (http://ceur-ws.org/Vol-2068/esida3.pdf)	Christopher Harris	Text	Summarization/Topic Analysis/ Entity Extraction Sentiment Analysis	Clustering/Classification/Categorization Comparison Overview	2018
ID5	Visualizing Nonlinear Narratives with Story Curves (https://doi.org/10.1109/TVCG.2017.2744118)	Nam Wook Kim, Benjamin Bach, Hyejin Im, Sasha Schriber, Markus	Text	Summarization/Topic Analysis/ Entity Extraction Discourse Analysis Sentiment Analysis Event Analysis Trend Analysis/Pattern Analysis	Region of Interest Clustering/Classification/Categorization Comparison Overview Navigation/Exploration	2018

				Gross, and Hanspeter Pfister.				
ID6	DoSVis: Document Stance Visualization (https://doi.org/10.5220/0006539101680175)			Kostiantyn Kucher, Carita Paradis, Andreas Kerren	Text Summarization/Topic Analysis/ Entity Extraction		Clustering/Classification/Categorization Comparison Overview Navigation/Exploration	2018
ID7	Organic Visualization of Document Evolution (https://doi.org/10.1145/3172944.3173004)			Ignacio Perez-Messina, Claudio Gutierrez, Eduardo Graells-Garrido	Event Analysis Translation/Text Analysis	Alignment	Comparison Overview Navigation/Exploration	2018

ID8	Exploring Emotions in Online Movie Reviews for Online Browsing. (https://doi.org/10.1145/3030024.3040982)	Nadeem Bader, Osnat Mokryn, Joel Lanir	Sentiment Analysis Relation/Connection Analysis	Clustering/Classification/Categorization Comparison Overview	2017
ID9	Diggersdiaries: Using Text Analysis to Support Exploration and Reading in a Large Document Collection. (https://doi.org/10.2312/eurp.20171156) Restrito.	Jaume Nualart Vilaplana and Mario Pérez-Montoro	Text Analysis/ Entity Extraction Summarization/Topic	Clustering/Classification/Categorization Comparison Overview Navigation/Exploration	2017
ID10	Interactive Ambiguity Resolution of Named Entities in Fictional Literature. (https://doi.org/10.1111/cgf.13179)	Florian Stoffel, Wolfgang Jentner, Michael Behrisch, Johannes Fuchs,	Text Analysis/ Entity Extraction Sentiment Analysis Relation/Connection Analysis Summarization/Topic	Clustering/Classification/Categorization Comparison Overview Navigation/Exploration	2017

			Daniel Keim.				
ID11	Task-Driven Comparison of Topic Models (https://doi.org/10.1109/TVCG.2015.2467618)	Eric Alexander, Michael Gleicher	Text Summarization/Topic Analysis/ Entity Extraction Relation/Connection Analysis Translation/Text Analysis	Alignment	Clustering/Classification/Categorization Comparison Overview Navigation/Exploration	2016	
ID12	Visualization of Latin Textual Variants using a Pixel-Based Text Analysis Tool. (https://doi.org/10.2312/eurova.2016119) Restrito	Bharathi Asokarajan, Ronak Etemadpour, June Abbas, Sam Huskey, Chris Weaver.	Lexical/Syntactical Analysis Translation/Text Analysis	Alignment	Clustering/Classification/Categorization Comparison Overview Navigation/Exploration	2016	
ID13	Towards Automatic Argument Extraction and Visualization in a Deliberative Model of Online Consultations for Local Governments	Robert Bembenik	Discourse Analysis Stance Analysis Sentiment Analysis Relation/Connection Analysis		Clustering/Classification/Categorization Comparison Overview	2016	

	(https://doi.org/10.1007/978-3-319-44039-2_6)		Piotr Andruszkiewicz				
ID14	Typographic Sets: Labelled Set Elements with Font Attributes (http://ceur-ws.org/Vol-1655/brath.pdf)	Richard Brath Ebad Banissi	Sentiment Analysis Relation/Connection Analysis	Clustering/Classification/Categorization Comparison Overview Navigation/Exploration	2016		
ID15	TimeLineCurator: Interactive Authoring of Visual Timelines from Unstructured Text. (https://doi.org/10.1109/TVCG.2015.2467531)	Johanna Fulda, Matthew Brehmer, Tamara Munzner.	Text Summarization/Topic Analysis/ Entity Extraction Trend Analysis/Pattern Analysis Relation/Connection Analysis	Clustering/Classification/Categorization Comparison Overview Navigation/Exploration Uncertainty Tackling	2016		
ID16	Speculative Practices: Utilizing InfoVis to Explore Untapped Literary Collections.	Uta Hinrichs, Stefania Forlini,	Text Summarization/Topic Analysis/ Entity Extraction Trend Analysis/Pattern Analysis	Clustering/Classification/Categorization Comparison	2016		

	(https://doi.org/10.1109/TVCG.2015.2467452)	Bridget Moynihan	Relation/Connection Analysis	Overview Navigation/Exploration	
ID17	Textual Affect Communication and Evocation Using Abstract Generative Visuals. (https://doi.org/10.1109/THMS.2015.2504081)	Uros Krcadinac, Jelena Jovanovic, Vladan Devedzic, Philippe Pasquier.	Sentiment Analysis	Clustering/Classification/Comparison Monitoring	2016
ID18	Visual analysis of online social media to open up the investigation of stance phenomena. (https://doi.org/10.1177/1473871615575079)	Kostiantyn Kucher, Teri Schamp-Bjerede, Andreas Kerren, Carita Paradis,	Text Summarization/Topic Analysis/ Entity Extraction Sentiment Analysis Stance Analysis Trend Analysis/Pattern Analysis Region of Interest	Clustering/Classification/Comparison Overview Navigation/Exploration	2016

			Magnus Sahlgren					
ID19	InVITe - Towards Intuitive Visualization of Iterations over Text. (https://doi.org/10.2312/eurp.2016114 1) Restrito	Houda Lamqaddam Jan Aerts		Translation/Text Analysis	Alignment	Comparison Overview Navigation/Explorati on	2016	
ID20	Poemage: Visualizing the Sonic Topology of a Poem. (https://doi.org/10.1109/TVCG.2015.2467811)	Nina McCurdy, Julie Lein, Katharine Coles, Miriah Meyer		Lexical/Syntactical Analysis Translation/Text Analysis Region of Interest	Alignment	Clustering/Classifica tion/Categorization Comparison Overview	2016	
ID21	Supporting Collaborative Political Decision Making: An Interactive Policy Process Visualization System. (https://doi.org/10.1145/2968220.2968223)	Tobias Ruppert, Andreas Bannach, Jürgen Bernard,		Sentiment Analysis Event Analysis		Clustering/Classifica tion/Categorization Comparison Overview Navigation/Explorati on	2016	

		Hendrik Lücke-Tieke, Alex Ulmer, Jörn Kohlhammer.			
ID22	Guidelines for Effective Usage of Text Highlighting Techniques. (https://doi.org/10.1109/TVCG.2015.2467759)	Hendrik Strobelt, Daniela Oelke, Bum Chul Kwon, Tobias Schreck, Hanspeter Pfister.	Region of Interest		2016
ID23	TextDNA: Visualizing Word Usage with Configurable Colorfields. (https://doi.org/10.1111/cgf.12918)	Danielle Albers Szafir, Deidre Stuffer,	Trend Analysis/Pattern Analysis Lexical/Syntactical Analysis Region of Interest	Comparison Overview Navigation/Explorati on	2016

		Yusef Sohail, Michael Gleicher				
ID24	A Guided Tour of Literature Review: Facilitating Academic Paper Reading with Narrative Visualization. (https://doi.org/10.1145/2968220.2968242)	Yun Wang, Dongyu Liu, Huamin Qu, Qiong Luo, Xiaojuan Ma	Sentiment Analysis Trend Analysis/Pattern Analysis Relation/Connection Analysis	Clustering/Classification/Categorization Comparison Overview Navigation/Exploration	2016	
ID25	Using Text in Visualizations for Micro/Macro Readings. (http://vialab.science.uoit.ca/textvis2015/papers/Brath-textvis2015.pdf)	Richard Brath Ebad Banissi.	Text Summarization/Topic Analysis/ Entity Extraction Region of Interest	Clustering/Classification/Categorization	2015	
ID26	Visualization System for Analyzing User Opinion. (https://doi.org/10.1109/SII.2015.7405055)	Yu-Sheng Chen, Lieu-Hen Chen, Toru Yamaguchi,	Text Summarization/Topic Analysis/ Entity Extraction Discourse Analysis Sentiment Analysis	Clustering/Classification/Categorization Comparison Overview Navigation/Exploration	2015	

		Yasufumi Takama.					
ID27	Morphable Word Clouds for Time-Varying Text Data Visualization. (https://doi.org/10.1109/TVCG.2015.2440241)	Ming-Te Chi, Shih-Syun Lin, Shiang-Yi Chen, Chao-Hung Lin, Tong-Yee Lee.	Text	Summarization/Topic Analysis/ Entity Extraction Trend Analysis/Pattern Analysis	Comparison Overview	2015	
ID28	uRank: Visual Analytics Approach for Search Result Exploration. (https://doi.org/10.1109/VAST.2015.7347686)	Cecilia di Sciascio, Vedran Sabol, Eduardo Veas	Text	Summarization/Topic Analysis/ Entity Extraction Lexical/Syntactical Analysis	Clustering/Classification/Categorization Comparison Overview Navigation/Exploration	2015	
ID29	Exploratory Text Analysis using Lexical Episode Plots.	Valentin Gold,	Text	Summarization/Topic Analysis/ Entity Extraction	Clustering/Classification/Categorization	2015	

	(https://doi.org/10.2312/eurovisshort.20151130)	Christian Rohrdantz Mennatallah El-Assady.	Sentiment Analysis Trend Analysis/Pattern Analysis	Overview Navigation/Exploration	
ID30	Visual Assessment of Alleged Plagiarism Cases. (https://doi.org/10.1111/cgf.12618)	Patrick Riehmann, Martin Potthast, Benno Stein, Bernd Froehlich	Relation/Connection Translation/Text Alignment Analysis Region of Interest	Clustering/Classification/Categorization Comparison Overview Navigation/Exploration	2015
ID31	The Stor-e-Motion Visualization for Topic Evolution Tracking in Text Data Streams. (https://doi.org/10.5220/0005292900290039)	Andreas Weiler, Michael Grossniklaus, Marc H. Scholl.	Text Summarization/Topic Analysis/ Entity Extraction Sentiment Analysis Event Analysis	Clustering/Classification/Categorization Overview Monitoring	2015
ID32	Visual Exploratory Tool for Storyline Generation.	Erk Ediz Akyiğit,	Text Summarization/Topic Analysis/ Entity Extraction	Clustering/Classification/Categorization	2014

	(https://doi.org/10.1109/VAST.2014.7042497)	Tuğkan Cengiz, Onur Burak Yıldırım, Selim Balcisoy	Relation/Connection	Comparison Overview Navigation/Exploration	
ID33	Using Font Attributes in Knowledge Maps and Information Retrieval. (http://ceur-ws.org/Vol-1311/paper3.pdf)	Richard Brath Ebad Banissi.	Lexical/Syntactical Analysis	Overview Navigation/Exploration	2014
ID34	V1 in Icelandic: A multifactorial visualization of historical data. (http://webservices.eurac.edu/vislr.html) Não disponível	Miriam Butt, Tina Bögel, Kristina Kotcheva, Christin Schätzle, Christian Rohrdantz,	Lexical/Syntactical Analysis	Comparison Overview Navigation/Exploration	2014

		Dominik Sacha, Nicole Dehe, Daniel A. Keim				
ID35	Linea: Tailoring Timelines by Visual Exploration of Temporal Text. (https://doi.org/10.1109/VAST.2014.7042513)	Tiago Etienne, Paulo Pagliosa, Luis Gustavo Nonato	Text Summarization/Topic Analysis/ Entity Extraction Event Analysis Trend Analysis/Pattern Analysis Region of Interest	Comparison Overview Navigation/Explorati on		2014
ID36	Word Cloud Explorer: Text Analytics Based on Word Clouds. (https://doi.org/10.1109/HICSS.2014.231)	Florian Heimerl, Steffen Lohmann, Simon Lange, Thomas Ertl.	Text Summarization/Topic Analysis/ Entity Extraction Lexical/Syntactical Analysis	Clustering/Classifica tion/Categorization Comparison Overview Navigation/Explorati on		2014
ID37	Lexichrome: Examining Word-Color Associations with Visualization.	Chris K. Kim Christopher Collins	Trend Analysis/Pattern Analysis Lexical/Syntactical Analysis	Clustering/Classifica tion/Categorization		2014

	(http://vialab.dcuoit.net/wordpress/wp-content/papercite-data/pdf/Kim2014a.pdf)						
ID38	VarifocalReader – In-Depth Visual Analysis of Large Text Documents. (https://doi.org/10.1109/TVCG.2014.2346677)	Steffen Koch, Markus John, Michael Wörner, Andreas Müller, Thomas Ertl.	Text Summarization/Topic Analysis/ Entity Extraction Relation/Connection Translation/Text Analysis Region of Interest	Alignment	Clustering/Classification/Categorization Comparison Overview Navigation/Exploration	2014	
ID39	Intelligent Interface for Textual Attitude Analysis. (https://doi.org/10.1145/2535912)	Alena Neviarouskaya, Masaki Aono, Helmut Prendinger, Mitsuru Ishizuka	Stance Analysis Sentiment Analysis Trend Analysis/Pattern Analysis		Clustering/Classification/Categorization Comparison Overview	2014	

ID40	Text Variation Explorer: Towards interactive visualization tools for corpus linguistics. (https://doi.org/10.1075/ijcl.19.3.05sii)	Harri Siirtola, Tanja Säily, Terttu Nevalainen, Kari-Jouko Räihä	Text Summarization/Topic Analysis/ Entity Extraction Trend Analysis/Pattern Analysis Lexical/Syntactical Analysis Region of Interest	Clustering/Classification/Categorization Comparison Overview Navigation/Exploration	2014
ID41	Semantize: visualizing the sentiment of individual document. (https://doi.org/10.1145/2598153.2600056)	Alan J. Wecker, Joel Lanir, Osnat Mokryn, Einat Minkov, Tsvi Kuflik.	Sentiment Analysis	Clustering/Classification/Categorization Navigation/Exploration	2014
ID42	Rule-based Visual Mappings – with a Case Study on Poetry Visualization. (https://doi.org/10.1111/cgf.12125)	Alfie Abdul-Rahman, Julie Lein, Katharine Coles,	Lexical/Syntactical Analysis Relation/Connection	Comparison Navigation/Exploration	2013

		Eamonn Maguire, Miriah Meyer, Martin Wynne, Chris R. Johnson, Anne Trefethen, Min Chen			
ID43	Visualizing patterns of appraisal in texts and corpora. (https://doi.org/10.1515/text-2013-0031)	Bandar Alhumaidi A. Almutairi	Discourse Analysis Stance Analysis Sentiment Analysis Trend Analysis/Pattern Analysis	Clustering/Classification/Categorization Comparison Overview	2013
ID44	Texty, a visualization tool to aid selection of texts from search outputs. (http://InformationR.net/ir/18-2/paper581.html)	Jaume Nualart and Mario Pérez-Montoro	Text Summarization/Topic Analysis/ Entity Extraction Relation/Connection Trend Analysis/Pattern Analysis	Clustering/Classification/Categorization Comparison Overview	2013

			Region of Interest	Navigation/Explorati on	
ID45	Fingerprint Matrices: Uncovering the dynamics of social networks in prose literature. (https://doi.org/10.1111/cgf.12124)	Daniela Oelke, Dimitrios Kokkinakis, Daniel A. Keim.	Text Summarization/Topic Analysis/ Entity Extraction	Comparison Overview	2013
ID46	Conceptual Recurrence Plots: Revealing Patterns in Human Discourse. (https://doi.org/10.1109/TVCG.2011.100)	Daniel Angus, Andrew Smith, Janet Wiles.	Discourse Analysis Region of Interest	Comparison Overview Navigation/Explorati on	2012
ID47	Termite: Visualization Techniques for Assessing Textual Topic Models. (https://doi.org/10.1145/2254556.2254572)	Jason Chuang, Christopher D. Manning, Jeffrey Heer.	Text Summarization/Topic Analysis/ Entity Extraction	Clustering/Classifica tion/Categorization Comparison Overview Navigation/Explorati on	2012

ID48	TopicNets: Visual Analysis of Large Text Corpora with Topic Modeling. (https://doi.org/10.1145/2089094.2089099)	Brynjar Gretarsson, John O'Donovan, Svetlin Bostandjiev, Tobias Höllerer, Arthur Asuncion, David Newman, Padhraic Smyth.	Text Summarization/Topic Analysis/ Entity Extraction Relation/Connection	Clustering/Classification/Categorization Overview	2012
ID49	From once upon a time to happily ever after: Tracking emotions in mail and books. (https://doi.org/10.1016/j.dss.2012.05.03)	Saif M. Mohammad	Sentiment Analysis Trend Analysis/Pattern Analysis Lexical/Syntactical Analysis	Clustering/Classification/Categorization Comparison Overview	2012

	Não disponível			Navigation/Explorati on	
ID50	WORDGRAPH: Keyword-in-Context Visualization for NETSPEAK's Wildcard Search. (https://doi.org/10.1109/TVCG.2012.96)	Patrick Riehmann, Henning Gruendl, Martin Potthast, Martin Trenkmann, Benno Stein, Bernd Froehlich.	Relation/Connection	Comparison Navigation/Explorati on	2012
ID51	Document Thumbnails with Variable Text Scaling. (https://doi.org/10.1111/j.1467-8659.2012.03109.x)	Andreas Stoffel, Hendrik Strobel, Oliver Deussen,	Region of Interest	Overview	2012

			Daniel A. Keim				
ID52	Exploring Collections of Tagged Text for Literary Scholarship. (https://doi.org/10.1111/j.1467-8659.2011.01922.x)	Text	Michael Correll, Michael Witmore, Michael Gleicher	Trend Analysis/Pattern Analysis Lexical/Syntactical Analysis Region of Interest	Overview		2011
ID53	xLDD: Extended Linguistic Dependency Diagrams. (https://doi.org/10.1109/IV.2011.42)	Text	Chris Culy, Verena Lyding, Henrik Dittmann	Summarization/Topic Analysis/ Entity Extraction Lexical/Syntactical Analysis	Comparison Overview		2011
ID54	Sampling latent emotions and topics in a hierarchical Bayesian network. (https://doi.org/10.1109/NLPKE.2011.6138166)	Text	Xin Kang Fuji Ren	Summarization/Topic Analysis/ Entity Extraction Sentiment Analysis Lexical/Syntactical Analysis Relation/Connection	Clustering/Classification/ Categorization Comparison Overview		2011

ID55	Visualising a Text with a Tree Cloud. (https://doi.org/10.1007/978-3-642-10745-0_61)	Philippe Gambette Jean Véronis	Text Analysis/ Entity Extraction	Summarization/Topic Entity Extraction	Clustering/Classification/ Categorization Overview	2010
ID56	Visualizing the Text of Philip Pullman's Trilogy "His Dark Materials". (https://doi.org/10.1145/1868914.1869023)	Tim Regan Linda Becker	Lexical/Syntactical Relation/Connection		Comparison Overview	2010
ID57	Bubble Sets: Revealing Set Relations with Isocontours over Existing Visualizations. (https://doi.org/10.1109/TVCG.2009.122)	Christopher Collins, Gerald Penn, Sheelagh Carpendale	Relation/Connection Translation/Text Analysis	Alignment	Clustering/Classification/ Categorization Uncertainty Tackling	2009
ID58	DocuBurst: Visualizing Document Content using Language Structure. (https://doi.org/10.1111/j.1467-8659.2009.01439.x)	Christopher Collins, Sheelagh Carpendale, Gerald Penn	Text Analysis/ Lexical/Syntactical Relation/Connection	Summarization/Topic Entity Extraction Analysis	Clustering/Classification/ Categorization Comparison Overview	2009

					Navigation/Explorati on	
ID59	Mapping Text with Phrase Nets. (https://doi.org/10.1109/TVCG.2009.165)	Frank van Ham, Martin Wattenberg, Fernanda B. Viégas	Lexical/Syntactical Analysis Relation/Connection		Overview	2009
ID60	The study on text emotional orientation based on a three-dimensional emotion space model. (https://doi.org/10.1109/NLPKE.2009.5313815)	Huana Li Fuji Ren	Sentiment Analysis		Clustering/Classifica tion/Categorization Overview	2009
ID61	Document Visualization Based on Semantic Graphs. (https://doi.org/10.1109/IV.2009.57)	Delia Rusu, Blaž Fortuna, Dunja Mladenić, Marko Grobelnik,	Text Summarization/Topic Analysis/ Entity Extraction Relation/Connection Region of Interest		Clustering/Classifica tion/Categorization Comparison Overview	2009

		Ruben Sipoš				
ID62	The Word Tree, an Interactive Visual Concordance. (https://doi.org/10.1109/TVCG.2008.172)	Martin Wattenberg Fernanda B. Viégas	Lexical/Syntactical Analysis Region of Interest			2008
ID63	Categorization and analysis of text in computer mediated communication archives using visualization. (https://doi.org/10.1145/1255175.1255178)	Ahmed Abbasi Hsinchun Chen	Text Analysis/ Entity Extraction Sentiment Analysis Trend Analysis/Pattern Analysis	Summarization/Topic Clustering/Classification/Categorization Comparison Overview		2007
ID64	Narrative based Topic Visualization for Chronological Data. (https://doi.org/10.1109/IV.2007.80)	Mina Akaishi, Ken Satoh, Yoshikiyo Kato, Koichi Hori	Text Analysis/ Entity Extraction Relation/Connection Region of Interest	Summarization/Topic Clustering/Classification/Categorization Comparison Overview		2007
ID65	Visualization of Uncertainty in Lattices to Support Decision-Making.	Christopher Collins, Sheelagh	Relation/Connection Translation/Text Analysis	Alignment Comparison Overview Uncertainty Tackling		2007

	(https://doi.org/10.2312/VisSym/EuroVis07/051-058) Indisponível	Carpendale, Gerald Penn					
ID66	Sequential Document Visualization. (https://doi.org/10.1109/TVCG.2007.70592)	Yi Mao, Joshua V. Dillon, and Guy Lebanon	Text	Summarization/Topic Analysis/ Entity Extraction Trend Analysis/Pattern Analysis Relation/Connection	Comparison Overview	2007	
ID67	Literature Fingerprinting: A New Method for Visual Literary Analysis. (https://doi.org/10.1109/VAST.2007.4389004)	Daniel A. Keim and Daniela Oelke	Text	Summarization/Topic Analysis/ Entity Extraction	Comparison Overview	2007	
ID68	Text Visualization – What Colors Tell About a Text. (https://doi.org/10.1109/IV.2007.108)	Wibke Weber	Text	Summarization/Topic Analysis/ Entity Extraction Lexical/Syntactical Analysis	Clustering/Classification/Categorization Comparison Overview	2007	

ID69	Scentindex: Conceptually Reorganizing Subject Indexes for Reading. (https://doi.org/10.1109/VAST.2006.261418)	Ed H. Chi, Lichan Hong, Julie Heiser, Stuart K. Card.	Text Summarization/Topic Analysis/ Entity Extraction Relation/Connection	Clustering/Classification/Categorization Overview Navigation/Exploration	2006
ID70	Visualizing the Performance of Computational Linguistics Algorithms. (https://doi.org/10.1109/VAST.2006.261417)	Stephen G. Eick, Justin Mauger, and Alan Ratner	Text Summarization/Topic Analysis/ Entity Extraction Lexical/Syntactical Analysis Relation/Connection	Clustering/Classification/Categorization Comparison Overview Navigation/Exploration	2006
ID71	User-directed sentiment analysis: visualizing the affective content of documents. (http://dl.acm.org/citation.cfm?id=1654645)	Michelle L. Gregory, Nancy Chinchor, Paul Whitney, Richard Carter,	Sentiment Analysis	Clustering/Classification/Categorization Comparison Overview	2006

		Elizabeth Hetzler, Alan Turner.				
ID72	A Corpus Study of Canned Letters: Mining the Latent Rhetorical Proficiencies Marketed to Writers-in-a-Hurry and Non-Writers. (https://doi.org/10.1109/TPC.2006.880743)	David Kaufer Suguru Ishizaki	Text Analysis/ Entity Extraction Lexical/Syntactical Analysis	Summarization/Topic Entity Extraction Analysis	Comparison Overview Navigation/Exploration	2006
ID73	Exploring erotics in Emily Dickinson's correspondence with text mining and visual interfaces. (https://doi.org/10.1145/1141753.1141781)	Catherine Plaisant, James Rose, Bei Yu, Loretta Auvil, Matthew G. Kirschenbaum, Martha Nell Smith,	Text Analysis/ Entity Extraction Relation/Connection	Summarization/Topic Entity Extraction Relation/Connection	Clustering/Classification/Categorization Comparison Overview Navigation/Exploration	2006

		Tanya Clement, Greg Lord				
ID74	Visualizing the meaning of texts. (https://doi.org/10.1109/IV.2005.153)	Wai K. Yeap, Paul Reedy, Kyongho Min, and Hilda Ho.	Text Analysis/ Lexical/Syntactical Relation/Connection	Summarization/Topic Entity Extraction Analysis	Clustering/Classifica tion/Categorization Comparison Overview Navigation/Explorati on	2005
ID75	Visualizing the affective structure of a text document. (https://doi.org/10.1145/765891.765961)	Hugo Liu, Ted Selker, Henry Lieberman	Sentiment Analysis		Clustering/Classifica tion/Categorization Comparison Overview Navigation/Explorati on	2003
ID76	TextArc: Showing Word Frequency and Distribution in Text. (http://www.textarc.org) Não disponível	W. Bradford Paley	Text Analysis/ Lexical/Syntactical Relation/Connection	Summarization/Topic Entity Extraction Analysis	Overview	2002

ID77	Arc diagrams: visualizing structure in strings. (https://doi.org/10.1109/INFVIS.2002.1173155)	Martin Wattenberg.	Trend Analysis/Pattern Analysis Relation/Connection	Comparison Overview	2002
ID78	Affect analysis of text using fuzzy semantic typing. (https://doi.org/10.1109/91.940962)	Pero Subasic Alison Huettner.	Sentiment Analysis Relation/Connection	Clustering/Classification/Categorization Comparison Navigation/Exploration	2001

Apêndice II. Artigos de visualização na mineração de texto na ER

NO.	TITULO	ATOR(ES)	TAREFA(S) ANALÍTICA(S)	CATEGORIA(S) VISUALIZAÇÃO	DE ANO
A1.	Perspectives on User Story Based Visual Transformations.	Yves Wautelet, Samedi Heng, Manuel Kolp.	Rotulado Resumo	Modelo Caso de Uso	2017

A2.	ReCVisu: A tool for clustering-based visual exploration of requirements.	Sandeep Reddivari, Zhangji Chen, Nan Niu	Agrupamento Resumo	Correlation	2012
A3.	Rigi—Na environment for software reverse engineering, exploration, visualization, and redocumentation.	Holger M. Kienle, Hausi A.Müller	Agrupamento Rotulado	Flow	2010
A4.	Visualizing requirements: A three-dimensional pyramid representation.	Deepti Savio, P.C. Anitha, Parameshwar P. Iyer	Manual refinements	3D	2011
A5.	Improved representation of traceability links in requirements engineering knowledge using Sunburst and Netmap visualizations.	Thorsten Merten, Daniela Jüppner, Alexander Delater		<i>Part of Whole</i> <i>Flow</i>	2011
A6.	Requirements traceability matrix: automatic generation and visualization.	André Di Thommazo, Gabriel Malimpensa, Thiago Ribeiro de Oliveira, Guilherme Olivatto, Sandra C. P. F. Fabbri	Similaridade	<i>Flow</i>	2012
A7.	Visualizing requirements in distributed system development.	Deepti Savio, PC Anitha,	Agrupamento	3D	2012

		Arpith Patil, Oliver Creighton			
A8.	Automated extraction and visualization of quality concerns from requirements specifications.	Mona Rahimi, Mehdi Mirakhorli, Jane Cleland-Huang	Rotulado Agrupamento	<i>Flow</i>	2014
A9.	Mining Requirements Knowledge from Collections of Domain Documents.	Xiaoli Lian, Mona Rahimi, Jane Cleland-Huang, Li Zhang, Remo Ferrai, Michael Smith	Realçado	Uso de cor no destaque do texto	2016
A10.	Let's Hear it from RETTA: A Requirements Elicitation Tool for Traffic Management Systems.	Mohammad Noaen, Zahra Shakeri Hossein Abad, Behrouz Homayoun Far	Rotulado	Text	2017
A11.	ECrits — Visualizing Support Ticket Escalation Risk.	Lloyd Montgomery, Emma Reading, Daniela Damian	Machine Learning	Etiquetas de texto.	2017
A12.	ELICA: An Automated Tool for Dynamic Extraction of Requirements Relevant Information.	Zahra Shakeri Hossein Abad Vincenzo Gervasi,	Realçado Rotulado	Uso de cor no destaque do texto	2018

		Didar Zowghi, Ken Barker	Resumo		
A13.	The Interactive Narrator Tool: Effective Requirements Exploration and Discussion through Visualization.	Govert-Jan Slob, Fabiano Dalpiaz, Sjaak Brinkkemper, Garm Lucassen	Realçado Rotulado	<i>Flow</i>	2018
A14.	EA-Miner: A Tool for Automating Aspect-oriented Requirements Identification.	A. Sampaio, R. Chitchyan, A. Rashid, P. Rayson	Rotulado	Text	2005

Apêndice III. Conteúdos de ajuda para avaliação

A mineração de texto tem vindo a aumentar o seu uso (KUCHER, KERREN, 2015) já que permite localizar e extrair as informações mais importantes e essenciais dos documentos, bem como conhecimento implícito e oculto (ZAYAS, MEDEROS, 2007). Com essa premissa a mineração influencia nas atividades de elicitação pois contribui na aquisição de conhecimento automaticamente, para a construção de um sistema útil e de qualidade. A mineração para a Engenharia de Requisitos pode:

- Usar o Agrupamento (*Clustering*) na exploração de tópicos relevantes dentro de uma coleção de documentos textuais e assim, melhorar a exploração dos requisitos [2]. Agrupar conceitos na engenharia reversa para gerar artefato [3] e refinar a rastreabilidade entre requisitos [4,7,8].
- Outra técnica como Realçado (*Highlighted*) é empregada para destacar trechos de textos com informações relacionada a requisitos [9] ou termos relevantes para a compreensão de domínio de aplicação [12].
- Usar cálculo de Similaridade (*Similarity*) para agrupar requisitos de acordo com sua descrição [2] além da detecção de dependência para rastreabilidade a partir de matrizes [6].
- O Rotulado (*Labeled*), é uma tarefa que não é usada solitariamente. Para algumas pesquisas [3,8,12,13], ela é exibida conjuntamente com ressaltado de um termo o para agrupar esses termos de acordo com um critério; salvo quando é usado para obter classificação do requisito [10], para facilitar a transformação para um diagrama de caso de uso [1] e para obter modelos de especificação mais refinados [14].

Esse resultado da mineração, geralmente texto sem formatação, pode ser difícil de entender assim como é difícil enxergar os relacionamentos que podem existir entre os dados. A ajuda da visualização permite entender e processar melhor os dados ou o texto mais rapidamente. Ter um suporte visual pode ser mais rápido e mais bem-sucedido para entender as informações tratada.

Apêndice IV. Avaliação aplicada

Pergunta de avaliação:

Como melhorar o processo de obtenção de conhecimento sobre visualização de texto minerado na Engenharia de Requisitos?

Motivação:

O apoio da mineração de texto para melhorar a obtenção de conhecimento e acelerar as tarefas de elicitación é considerada uma solução viável que vem aumentando. A realização da mineração de texto utilizando técnicas de visualização pode auxiliar na detecção de relacionamentos entre informações, especificamente na coleta de informações (fatos) a partir da automatização na leitura dos documentos.

Esta avaliação será uma maneira de ajudar a conhecer, de forma visual, sobre a mineração de texto para elicitación requisitos.

Dados de caracterização

Idade: ____

Escolaridade Grau: () básico () médio () superior () pós-graduação

Profissão: _____

Anos de experiência trabalhando com mineração de texto: ____

Objetivo: Medir a relação qualidade-tempo, para o engenheiro de requisitos, com a seleção de visualizações para apresentar o resultado da mineração de texto.

Atividade 1: Poderia procurar em 10 min pesquisas que incluam visualizações de texto minerado para Engenharia de Requisitos. Pode usar o método com que sinta comodidade. Das pesquisas encontradas identifique: Autor(es), tarefa de mineração executada e a imagem da visualização proposta.

Atividade 2: Para fins de informação, saiba que existe a ferramenta “*Survey of Text Visualization Techniques*” (em <http://textvis.lnu.se/>), que exhibe pesquisas de visualização para textos que foram minerados. Em 2 minutos pode explorar o seu funcionamento.

Atividade 3: Com a ideia da ferramenta citada, foi criada a "*MinViewREDL*" (em <http://35.247.212.133/>), que tem como objetivo organizar o conhecimento das visualizações de texto minerado produzidas para a Engenharia de Requisitos. Usando a *MinViewREDL* poderia procurar em 10 min pesquisas que incluam visualizações de texto minerado para Engenharia de Requisitos.

Atividade 4: Você nos ajudará a obter resultados para avaliarmos essa nova ferramenta criada ao completar esse breve questionário:

No	Perguntas	
1.	De modo geral, eu achei a ferramenta...	Muito ruim _(1)_(2)_(3)_(4)_(5) Muito boa
2.	Achei que o uso da ferramenta foi...	Muito difícil _(1)_(2)_(3)_(4)_(5) Muito fácil
3.	Achei que o objetivo que persegue a ferramenta foi ...	Muito ruim _(1)_(2)_(3)_(4)_(5) Muito bom
4.	Entender as classificações propostas pela ferramenta foi ...	Muito difícil _(1)_(2)_(3)_(4)_(5) Muito fácil
5.	Você encontrou o que procurava?	<input type="checkbox"/> Não <input type="checkbox"/> Sim <input type="checkbox"/> Mais ou menos
6.	Você achou a navegação fácil e intuitiva?	<input type="checkbox"/> Não <input type="checkbox"/> Sim <input type="checkbox"/> Outra
7.	Indique fatores positivos que encontrou na ferramenta.	
8.	Indique fatores negativos que encontrou na ferramenta.	
9.	Pode dar alguma sugestão de melhoria da ferramenta?	

Apêndice V. Estórias de Usuário

A seguir os requisitos propriamente ditos (r) usando uma modelagem com Estórias de usuário (*User-Stories*).

US-r01
Nome do requisito: Inserir categorias
Iteración: 1
Prioridad: Alta
Autor: Administrador do sistema
Descrição: <ol style="list-style-type: none"> 1. Eu como Administrador da biblioteca quero inserir uma nova categoria. 2. A categoria é mostrada na lista de categorias do sistema.
Validação: Visualizar a categoria na lista correspondente.

US-r02
Nome do requisito: Inserir tarefas analíticas
Iteración: 1
Prioridad: Alta
Autor: Administrador do sistema
Descrição: <ol style="list-style-type: none"> 1. Eu como Administrador da biblioteca quero inserir uma nova tarefa analítica. 2. A tarefa analítica é mostrada na lista de categorias do sistema.
Validação: Visualizar a tarefa analítica na lista correspondente.

US-r03
Nome do requisito: Inserir artigo
Iteración: 1
Prioridad: Alta
Autor: Administrador do sistema
Descrição: <ol style="list-style-type: none"> 1. Eu como Administrador da biblioteca quero inserir um novo artigo. 2. Ao inserir os dados de artigo tenho que especificar em qual categoria está classificado. 3. Ao inserir os dados de artigo tenho que especificar que tarefa analítica é executada. 4. O artigo é mostrado na tela principal da Biblioteca.
Validação: Visualizar o artigo na tela principal da Biblioteca.

US-r04
Nome do requisito: Visualizar artigo
Iteración: 1
Prioridad: Alta

Autor: Usuário
Descrição: <ol style="list-style-type: none"> 1. Eu como usuario aceso a Biblioteca e visualizo todos os artigos. 2. Os artigos existentes são mostrados na tela principal da Biblioteca.
Validação: Visualizar todos os artigos inseridos, na tela principal da Biblioteca.

US-r05
Nome do requisito: Filtrar artigos por categoria
Iteración: 2
Prioridad: Alta
Autor: Usuário
Descrição: <ol style="list-style-type: none"> 1. Eu como Usuário da biblioteca quero visualizar o(s) artigo(s) de uma categoria selecionada. 2. São mostrados todos os artigos que tem essa categoria.
Validação: Visualizar o(s) artigo(s) correspondente(s) a essa categoria na tela principal da Biblioteca.

US-r06
Nome do requisito: Filtrar artigos por tarefa analítica
Iteración: 2
Prioridad: Alta
Autor: Usuário
Descrição: <ol style="list-style-type: none"> 1. Eu como Usuário da biblioteca quero visualizar o(s) artigo(s) de uma tarefa analítica selecionada. 2. São mostrados todos os artigos que tem essa tarefa analítica.
Validação: Visualizar o(s) artigo(s) correspondente(s) a essa tarefa analítica na tela principal da Biblioteca.

US-r07
Nome do requisito: Procurar um artigo
Iteración: 2
Prioridad: Alta
Autor: Usuário
Descrição: <ol style="list-style-type: none"> 1. Eu como Usuário da biblioteca quero procurar artigo(s) introduzindo uma expressão de busca. 2. São mostrados todos os artigos que contém essa expressão no título ou nome dos autores.
Validação: Visualizar o(s) artigo(s), que contém essa expressão, na tela principal da Biblioteca.

US-r08
Nome do requisito: Visualizar detalhes do artigo
Iteración: 2
Prioridad: Alta
Autor: Usuário
Descrição: <ol style="list-style-type: none">1. Eu como Usuário da biblioteca quero visualizar detalhes de um artigo selecionado.2. São mostrados todos os dados do artigo.
Validação: Visualizar os dados do artigo na tela principal da Biblioteca.