

Revista Brasileira de Computação Aplicada, Novembro, 2020

DOI: 10.5335/rbca.v12i3.10389 Vol. 12, Nº 3, pp. 51-60

Homepage: seer.upf.br/index.php/rbca/index

#### ARTIGO ORIGINAL

# Descoberta de conhecimento em dados não estruturados: Uma análise em um chat de jogo online

# Knowledge Discovery in Unstructured Data: An analysis in a chat of online game

Marla Rosana Pereira Melo <sup>10,1</sup> and Diana Francisca Adamatti <sup>10,1</sup>

<sup>1</sup>Centro de Ciências Computacionais – Universidade Federal do Rio Grande (C3/FURG) \*marlamelo.sinfo@gmail.com; dianaada@gmail.com; ...

Recebido: 12/12/2019. Revisado: 10/06/2020. Aceito: 20/07/2020.

#### Resumo

Este trabalho aborda um estudo e a análise sobre a descoberta de conhecimento, a partir da mineração de dados textuais, aplicada no chat de um jogo massively multiplayer online (MMO). Para isso, utilizou-se a ferramenta Text Mining Suite, onde o processo de descoberta de conhecimento, além da mineração de textos, é baseado na criação de conceitos definidos por uma pessoa que possui algum entendimento acerca dos textos à serem analisados. Também como forma complementar a pesquisa, foi utilizada a ferramenta de mineração de textos Sobek, como suporte à Text Minig Suite para a criação de conceitos. Esse artigo propõe uma metodologia experimental para a aplicação da descoberta de conhecimento em dados textuais gerados a partir de conversas dos jogadores. Com essa metodologia, avaliamos os resultados de três testes. O primeiro teste envolveu a ferramenta Text Mining Suite para extrair conhecimento da base de dados textuais; no segundo teste foi utilizada a ferramenta Sobek para extrair os conceitos do texto; por fim, o último teste agregou as duas ferramentas. Uma análise quantitativa possibilitou a apresentação da qualidade dos resultados da mineração de textos. Observou-se uma melhora nos resultados, um ganho de conhecimento quando as ferramentas Text Mining Suite e Sobek foram utilizadas, em conjunto, para o processo da descoberta de conhecimento.

Palavras-Chave: Descoberta de conhecimento em texto; Jogo; Sobek; Text mining suíte

#### Abstract

This paper presents a study and an analysis on knowledge discovery, basing on textual data mining, applied in a chat of a massively multiplayer online game (MMO). In this study, the tool was used Text Mining Suite, where the process of knowledge discovery, besides the mining of texts, is based on the creation of concepts defined by a person who has some understanding about the texts to be analyzed. Also, as a complementary way of research, Sobek text mining tool was used as a support to the Text Minig Suite for the creation of concepts. This paper proposes an experimental methodology for the application of the knowledge discovery in textual data generated from the conversations of the players. With this methodology, we evaluated the results of three tests. The first test involved Text Mining Suite tool to extract knowledge from the textual database; the second one was used Sobek tool to extract the concepts from the text; finally, the last test added the two tools. A quantitative analysis made possible the presentation of the quality of the results of the mining of texts. We observed an improvement in the results, a gain of knowledge when the tools Text Mining Suite and Sobek were used together for the process of the knowledge discovery.

Keywords: Knowledge discovery in text; Games; Sobek; Text minig suíte.

### 1 Introdução

Os jogos que permitem a constante interação entre os jogadores do mundo todo em um chat do jogo geram um grande volume de dados. Com isso, existe oportunidade de analisar e extrair conhecimento útil destes dados com o uso de técnicas de aprendizado de máquina. A análise de dados armazenados em formato não estruturado pode ser considerada uma atividade mais complexa, se comparada à análise de dados estruturados, justamente pelo fato dos dados não possuírem a característica da não estruturação (Ambrozio and Mo-

Dessa maneira, este trabalho propõe extrair conhecimento da troca de mensagens do jogo online Mobile Strike<sup>1</sup>, e apresentar a análise dos resultados obtidos por meio da mineração dos dados textuais, de forma a descobrir as principais regras e associações realizadas pelos jogadores durante o jogo.

Para isso, as ferramentas de mineração de dados textuais: Sobek<sup>2</sup> e Text Mining Suite<sup>3</sup>, foram utilizadas para a análise da base de dados textual constituída através do chat do jogo. O resultado alcançado pela descoberta de conhecimento da coleção de textos foi obtido por três testes e análise dos dados, onde observou-se que a aplicação das duas ferramentas em conjunto concebeu resultados melhores.

A estruturação do texto está dividida na seguinte forma: A seção 1 apresenta o tema alvo da pesquisa e a descrição dos objetivos; a seção 2 apresenta o referencial teórico e aplicado deste trabalho, incluindo jogos Massively Multiplayer Online (MMO), descoberta de conhecimento em dados não estruturados baseado em mineração de textos, as ferramentas utilizadas para a realização deste trabalho e os trabalhos relacionados; a seção 4 apresenta a metodologia aplicada para o desenvolvimento deste trabalho e as ferramentas utilizadas para a descoberta de conhecimento; na seção 5 são apresentados os resultados obtidos e a análise destes; por fim, a seção 6 apresenta a conclusão, onde as principais descobertas da análise dos resultados são reforçadas, também apresenta sugestões para trabalhos futuros.

## Referencial Teórico e Aplicado

#### 2.1 Jogos Massively Multiplayer Online

Vivemos em uma realidade onde os jogos eletrônicos estão presentes na vida de jovens e adultos em diferentes contextos e formas através do computador, em consoles de mesa, consoles portáteis, celulares, tablets, redes sociais, sites de entretenimento, na educação, em treinamentos e em muitos outros (Barbat et al., 2014).

Para Novak (2010), os jogos podem melhorar as habilidades físicas e mentais, proporcionando um treinamento inesperado para certas profissões. Os jogadores podem ser motivados a jogar jogos para adquirir conhecimentos sobre conceitos, processos e estratégias específicas, também muitos jogadores jogam para escapar das tensões e dos desafios da vida real. Além disso, atualmente, nota-se que há uma expansão de jogadores que jogam profissionalmente.

Com o aumento da demanda, a produção de jogos inovadores e de alta qualidade está ganhando mais espaço e notoriedade. Um dos segmentos de jogos que está em constante evolução e crescimento na indústria de jogos é do tipo Massively Multiplayer Online (MMO).

Esses jogos podem ser jogados em consoles de mesa e portáteis, bem como em smartphones, tablets por meio de uma cópia do jogo instalada em equipamento local conectado à internet, ou ainda diretamente na rede, via browser streaming (Fleury et al., 2014).

Os jogos MMO são interessantes porque possibilitam a comunicação entre jogadores espalhados por todo o mundo, além disso, Novak (2010) menciona que esses jogos estimulam a cooperação, colaboração e participação em equipes com a formação de grupos para derrotar os adversários, bem como prestar ajuda e resolver problemas juntos.

#### 2.2 Descoberta de conhecimento em dados não estruturados

Basicamente, dados não estruturados são aqueles que possuem estrutura livre. Segundo Zambenedetti (2002), são dados que estão na forma de texto em linguagem natural. Uma enorme quantidade de informação existe na forma da linguagem natural, pelo fato do aumento da facilidade de produção de conteúdo em forma de textos, sobretudo em redes sociais, blogs, e-mails e aplicações de conversação em tempo real. No sentido de extrair alguma informação relevante desses dados não estruturados, o processo de descoberta de conhecimento em textos pode ser utilizado.

A descoberta de conhecimento em textos é definida por Loh et al. (2000) como a aplicação de técnicas e ferramentas computacionais com o objetivo de auxiliar na busca de conhecimento novo e útil em coleções textuais. Para Carvalho et al. (2011), quando os dados que estão sendo analisados se apresentarem na forma de textos, é usual se referir à sua análise como mineração de textos (MT).

O processo de mineração de textos é apresentado na Fig. 1. A etapa de coleta de textos é a primeira tarefa no processo de mineração de texto, ela tem a função de formar a base de dados textuais a ser utilizada em todo o processo (Schiessl, 2007). A etapa seguinte é referente ao pré-processamento, nessa etapa, são utilizadas as técnicas de recuperação de informação (RI) ou técnicas de extração da informação (EI) para transformar os textos em forma estruturada, ou seja, atributo-valor. De um modo geral, a etapa de pré-processamento tem a finalidade de aprimorar a qualidade da representação dos dados a serem utilizados e organizá-los. As etapas do pré-processamento visam ajustar os dados textuais, a ponto de poderem ser submetidos a algum algoritmo de indexação ou de mineração de dados (Soares, 2010).

As ferramentas Sobek e Text Mining Suite, realizam as seguintes técnicas de recuperação de informação na

<sup>1</sup>http://www.mobilestrikeapp.com

<sup>&</sup>lt;sup>2</sup>http://sobek.ufrgs.br/

<sup>&</sup>lt;sup>3</sup>http://intext.com.br/index.php

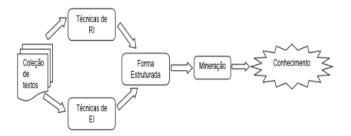


Figura 1: Processo de mineração de textos (Corrêa et al., 2003).

etapa de pré processamento da mineração de textos:

- Remoção de stopwords remoção das palavras irrelevantes;
- Processo de tokenização significa quebrar o texto em palavras;
- Indexação nesta etapa, ocorre a transformação dos dados para forma estruturada. A indexação do texto ocorre automaticamente em várias ferramentas quando os documentos são carregados (Rezende, 2003).

A próxima etapa refere-se a mineração, onde são aplicados técnicas e algoritmos adequados para tratar os dados e a partir disso, prosseguir para a etapa seguinte que consiste na analise dos resultados obtidos da mineração e na verificação do conhecimento útil e novo gerado da coleção de textual.

#### 2.3 Sobek

A ferramenta Sobek, conforme Klemann et al. (2009), utiliza um algoritmo específico, definido por Shanker em 2003, baseado em análises estatísticas dos textos e representação das informações extraídas em um modelo de grafo chamado de distância n-simples (nsimple distance). Os grafos apresentam informações referentes ao número absoluto e relativo de ocorrência com que as palavras ocorrem no texto (vértices) e relacionamentos (arestas) em determinado documento (Martins et al., 2010). Para Epstein and Reategui (2015), a ferramenta Sobek identifica os relacionamentos entre os conceitos baseando-se na proximidade em que as palavras estão uma da outra no texto e isto pode ser considerado como uma relação de Se-Então, uma relação ao temporal ou mesmo que os conceitos têm significados relativos um do outro. Um conceito pode estar relacionado com mais de um e, portanto, a ferramenta Sobek limita o número máximo de relações r = 7 para cada conceito. Dessa forma, a ferramenta consegue gerar as relações mais significativas entre os conceitos. A frequência com que um conceito ocorre no texto define o número de ligações que um conceito c pode ter. Dessa forma, os conceitos mais recorrentes (e, teoricamente, mais relevantes ao texto) possuem mais conexões que os demais conceitos e se destacam em posições centrais do grafo (Epstein and Reategui, 2015).

#### 2.4 Text Mining Suite

A ferramenta Text Mining Suite realiza o processo de descoberta de conhecimento através de algoritmos de mineração de textos (análise de frequência e associações entre conceitos) e baseia-se na criação de ontologias mediante conceitos definidos pelo utilizador da ferramenta. A ontologia é uma estrutura que descreve a existência ou específica de forma explícita e detalhada algo que, nós humanos, concebemos sobre o mundo (Sperotto, 2013). Ainda, de acordo com Sperotto (2013), o objetivo de toda ontologia é oferecer um caminho específico onde máquinas possam operar com bases de conhecimento que devem ser especificadas por especialistas humanos, para que os sistemas possam interoperar de acordo com regras ou fragmentos de conhecimento sobre um determinado domínio. Uma definição para conceitos, segundo Sowa et al. (2000), citado por Loh (2001), é a de que conceitos são expressos por linguagens (palavras e gramáticas), mas pertencem ao conhecimento extra-linguístico sobre o mundo.

Conforme o manual disponibilizado, o processo padrão de mineração dos textos para a utilização da ferramenta consiste nos seguintes passos:

- Pré-processamento: escolha do idioma, separação dos textos (se estiverem todos dentro de um mesmo documento), análise e representação interna dos textos:
- Criação da ontologia: Especialista define os conceitos e as regras para identificação dos conceitos nos textos;
- Mineração sobre os conceitos (Concept-based Text Mining): descoberta dos conceitos nos textos, análise estatística dos conceitos (frequência nos textos) e descoberta de associações entre conceitos (Data Mining).

#### 2.5 Trabalhos Relacionados

Neste contexto, no trabalho de Musabirov et al. (2015) foi abordado a mineração de textos aplicada ao jogo Castlot, um jogo MMO baseado em browser, onde combinou-se agentes qualitativos e métodos computacionais para mapear, através da coleta dos logs do chat do jogo, o conteúdo das perguntas dos jogadores iniciantes, com o objetivo de propor melhorias no desenvolvimento de tutoriais do jogo. Os autores afirmaram que foi possível detectar a experiência, adaptação de jogo dos jogadores.

Na pesquisa de Almeida et al. (2017), o processo de descoberta de conhecimento foi aplicado ao jogo Multiplayer Online Battle Arena (MOBA), um jogo do tipo eSport. Os autores objetivaram a previsão do rumo de uma partida com a aplicação de algoritmos de classificação para determinar o time com mais chance de vencer.

Já Thompson et al. (2017) realizaram uma análise de mensagens em jogos de StarCraft para extrair sentimentos dos textos. Para tal, utilizaram a ferramenta de extração de sentimentos SO-CAL (Semantic Orientation Calculator), modelo baseado em léxico. Os autores men-

cionam que a ferramenta SO-CAL contém cinco dicionários para adjetivos, substantivos, verbos, advérbios e intensificadores que carregam sentimentos, onde para cada palavra é atribuída uma pontuação de sentimento, classificadas em uma escala de -5 a 5. As palavras que não suportam sentimentos não são armazenadas no dicionário e dessa forma os autores adaptaram a ferramenta para o seu domínio de interesse. O modelo construído pelos autores para detecção de toxicidade e a classificação de sentimentos em jogos mostrou que métodos baseados no léxico são úteis para análise de sentimentos.

No trabalho de Wang and Goh (2020), a análise de texto em jogos on-line é utilizada para obter automaticamente componentes da experiência de jogo on-line. Na pesquisa, os autores analisaram e examinaram a importância relativa de cada componente para a satisfação do usuário, através da modelagem de tópicos para identificar componentes de jogos a partir de análises das avaliações dos jogos pelos jogadores. Os algoritmos de modelagem de tópicos são ferramentas estatísticas para analisar palavras em um texto e para descobrir os temas dominantes nele, de forma a obter uma organização do resumo de documentos. Também utilizaram a análise de sentimentos para extrair atitudes e emoções dos indivíduos em cada revisão. Neste caso a pontuação de sentimentos de cada revisão foi calculada combinando um dicionário estabelecido de sentimentos. Esta pesquisa revela que existem alguns componentes essenciais para os usuários para as avaliações de jogos.

Murnion et al. (2018) apresentam um sistema automático para coletar continuamente dados de bate-papo no jogo multijogador on-line: World of Tanks. Os autores mencionam que o método funciona através da raspagem na Web de um site de repetição de combinações, combinada às técnicas de Extração, Transformação e Carregamento (ETL) para criar um novo banco de dados de mensagens de bate-papo no jogo. O sistema, conjuntamente como a técnica de análise de sentimentos, foi aplicado para demostração da utilização dos dados para detecção automática de mensagens de bate-papo sobre cyberbullying.

### 3 Metodologia

A metodologia deste trabalho objetiva-se em gerar conhecimento a partir de conversas em um jogo mediante a aplicação da mineração de textos. Inicialmente, a pesquisa explorou artigos publicados em bases acadêmicas (Google Scholar e Science Direct), com o objetivo de identificar e analisar estudos sobre descoberta de conhecimento, com um foco em jogos e mineração de textos.

Neste estudo, os dados analisados são provenientes do chat do jogo Mobile Strike, que consiste em um jogo online de estratégia em tempo real, ou seja, não há partidas e tem como cenário a guerra moderna. No chat todos os jogadores podem trocar mensagens. O objetivo desse jogo é essencialmente o ganho de poder e a conquista do ponto de controle do estado no qual sua aliança (grupo de vários jogadores) está localizada.

A escolha do Mobile Strike deve-se ao fato também

da escolha da ferramenta Text Mining Suite, pois para realizar teste nesta ferramenta, há necessidade de conhecer bem conceitos do problema em questão para definir a ontologia necessária para sua execução. Assim, uma das autoras é jogadora do Mobile Strike e tinha condições de estruturar a ontologia necessária para os testes. No caso do Sobek, é uma ferramenta de fácil utilização e com execução via web, o que permitiu sua rápida integração na metodologia proposta.

Na Fig. 2 é ilustrada a metodologia proposta para a aplicação do processo de descoberta de conhecimento, com as ferramentas Text Mining Suite e Sobek.

O desenvolvimento do trabalho se dividiu nas seguintes etapas:

Etapa o: Coleta das conversas no chat do jogo Mobile Strike, manualmente, e armazenamento no formato de arquivo de texto para serem aplicadas nas ferramentas de mineração utilizadas neste trabalho. A coleta ocorreu no período de três meses, onde uma das autoras do trabalho fazia parte de um destes chats e coletou as informações realizadas pelo grupo, só que manteve o anonimato de todos os jogadores, mesmo que os mesmos, usualmente, joguem com nomes fictícios. Após a finalização da coleta dos textos, o arquivo, sem tratamento, constitui-se em um total de 8.081 linhas e 55.397 palavras.

Etapa 1: Preparação dos textos, onde realizou-se limpeza dos dados, dessa forma, foram eliminadas as palavras que não são relevantes para o processo de descoberta de conhecimento, pois apresentam pouca significância, como: "lol", "kkkk", "lmao", bem como figuras e nomes de jogadores. Depois deste processo de exclusão, a base de dados utilizada neste trabalho constituiu-se de um arquivo com 6.688 linhas e 50.181 e palavras.

Etapa 2: Processo de mineração que possui as subetapas 2A e 2b. Na primeira etapa 2A, foi utilizada a ferramenta de mineração de textos Text Mining Suite e derivou a etapa 3A. E na segunda etapa 2B, utilizou-se a ferramenta Sobek, que derivou as etapas 3B e 4.

Etapa 3A: Primeiro teste, onde obteve-se o primeiro resultado da mineração dos textos com a ferramenta text Minig Suite.

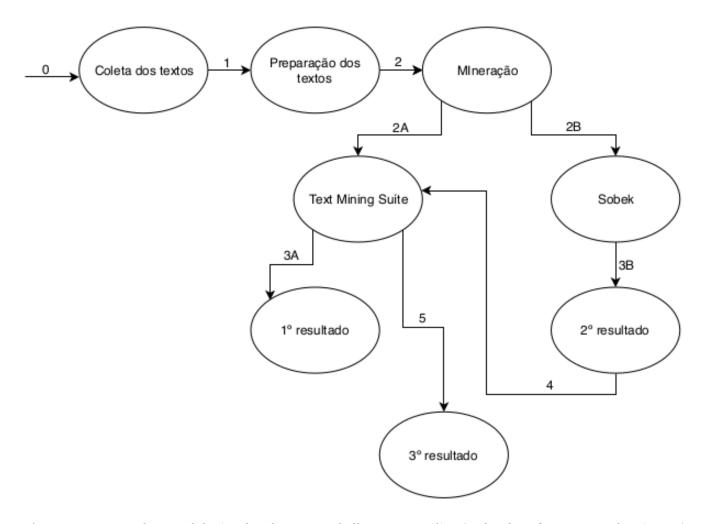
Etapa 3B: Segundo teste, onde obteve-se o resultado da mineração com a utilização da ferramenta Sobek.

Etapa 4: Térceiro teste que é dependente do resultado da etapa 3B, pois os conceitos gerados pela ferramenta Sobek foram inseridos como conceitos na ferramenta Text Mining Suite.

Etapa 5: Resultante da etapa 4, que com a utilização dos conceitos gerados pela ferrramenta Sobek como entrada para a ferramenta Text Mining Suite, gerou o terceiro resultado para análise.

#### 4 Resultados

Uma preparação dos textos foi necessária para não causar ruídos nos dados. Dessa forma, foi realizada uma limpeza nos dados com o auxílio de uma planilha ele-



**Figura 2:** Proposta de metodologia adotada neste trabalho, com a utilização das duas ferramentas de mineração de textos escolhidas.

trônica e um documento de texto, onde removeu-se os nomes de jogadores, *emoticons* e expressões de emoção.

A ferramenta Text Mining Suite não aceita um único arquivo de texto para realizar a mineração de texto, por esse motivo, as conversas foram proporcionalmente divididas em 100 arquivos de textos (número máximo de arquivos de textos permitido pela ferramenta). Isso é necessário pois a ferramenta é baseada na associação entre os conceitos presentes nos textos.

A identificação de um conceito no texto ocorre segundo as regras declaradas que caracterizam esse conceito, isto é, as palavras que definem o significado de um conceito. Portanto, se uma determinada palavra que caracteriza um assunto não estiver presente no texto, isso não significa que o texto não está se referindo a esse assunto.

Como a ferramenta baseia-se em conceitos para a descoberta de conhecimento, a definição destes foi concebida através do entendimento do jogo. Na Fig. 3 são apresentadas as regras criadas para a definição dos conceitos. O caractere \$ ao final de uma palavra é utilizado para indicar que outros caracteres podem aparecer ou

não. A adição do sinal de menos "-" antes da palavra indica que o conceito não existe.

	Conceitos	Regras do conceito
1	aliado	aliad\$, amig\$
2	aliança	aliança, colmeia,vage\$,hive\$, pods, pod
3	coletar	coleta\$, explora\$, recolhe\$, reuni\$, cultiv\$, telha, azulejo\$, agricultura, floresta\$, fazenda
4	família	familia\$, irmā, sub\$
5	guerra	guerra, armadilha\$, batalha\$, bat\$ - batepapo - batedeira, atac\$, atingi\$, hit\$, queim\$, blinda\$, escudo\$, rali, rally, zer\$
6	tropa	tropa\$, implantaç\$, t1, t2, t3, t4, tática, infantaria, artilharia, soldado\$, patrulha, morteiro, unidade de metralhadora, comando, veiculo\$, lança misseis, infantaria de foguete, elite, vbl, obus, demolições, delta, tanque, artilharia de foguete, homens

**Figura 3:** Conceitos definidos para a descoberta de conhecimento no jogo - Ontologia do jogo.

# 4.1 1º Resultado (3A) - Ferramenta Text Mining Suite

O primeiro resultado, refere-se a etapa 3A demostrado na Fig. 2 da metodologia. Com os 100 arquivos de textos e os 6 conceitos criados, descritos anteriormente, a ferramenta Text Mining (etapa 2A) gerou 15 regras associativas. O resultado apresenta as associações descobertas entre os conceitos.

Essas regras associativas são geradas a partir dos seguintes passos, presentes no manual da ferramenta text mining Suite:

- i. Identificação de conceitos nos textos conforme as regras definidas;
- ii. Extração do centroide contextual, onde é feita a análise de distribuição (frequência) dos conceitos nos textos;
- iii. Associações conceituais e geração de regras do tipo Se-Então. Este sub-processo identifica associações entre conceitos, mostrando a confiança (probabilidade condicional da associação) e o suporte (número de casos onde a associação aparece). As associações são apresentadas na forma de regras Se-Então.

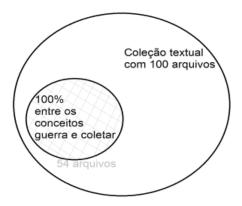
As associações que aparecem mais a esquerda representam a confiança da regra (probabilidade condicional). O número de textos em que as regras foram listadas representa o valor absoluto e valor percentual do suporte, isto é, se um conceito *c1* acontece nos textos então o conceito *c2* também ocorre.

Na Tabela 1 são apresentadas as associações geradas entre conceitos pela ferramenta Text Mining Suite.

**Tabela 1:** Associações entre conceitos no primeiro resultado

Confiança da Regra	Regras Associativas	Frequência
100%	aliança → guerra	43%
100%	$\operatorname{coletar}  o \operatorname{guerra}$	54%
100%	família  o guerra	22%
94.29%	aliado $ ightarrow$ guerra	33%
68.57%	aliado $ ightarrow$ coletar	24%
68.52%	$\operatorname{coletar}  o \operatorname{tropa}$	37%
67.44%	aliança $ ightarrow$ coletar	29%
63.64%	família  o tropa	14%
60.47%	aliança $ ightarrow$ tropa	26%
57.14%	aliado $\rightarrow$ aliança	20%
52.69%	$\mathbf{guerra}  o \mathbf{tropa}$	49%
51.43%	aliado $ ightarrow$ tropa	18%
29.63%	coleta  o família	16%
27.91%	aliança $ ightarrow$ família	12%
20%	aliado $\rightarrow$ família	07%

Na regra de associação: 100% dos textos que falam coleta também falam em guerra. Isso aconteceu em 54 textos (54% do total), que foi descoberta, o percentual encontrado representa que de todos os 100 arquivos da coleção textual, em 54 arquivos sempre que o conceito coletar estava presente no texto o conceito guerra estava também, como é exemplificado na Fig. 4.



**Figura 4:** Confiança da regra guerra  $\rightarrow$  coletar.

#### 4.2 2º Resultado (3B) - Sobek

O segundo resultado obtido, refere-se a etapa 3B demonstrado na Fig. 2 da metodologia. Para o processo de mineração de textos com ferramenta Sobek (etapa 2B), é necessário importar um único arquivo de texto ou copiar e colar o texto na interface da ferramenta. Na Fig. 5 são apresentados os conceitos gerados pela ferramenta Sobek.

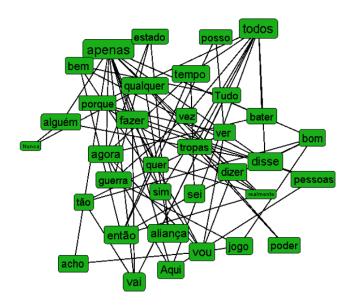


Figura 5: Grafo gerado pela ferramenta Sobek

Pode-se observar pela Fig. 5 que alguns conceitos extraídos não agregam conhecimento, como os termos: vai, acho, então, aqui, vou, realmente, bom, nunca, bem, apenas. Portanto, os conceitos considerados relevantes foram: aliança, estado, guerra, jogo, pessoas, poder, tempo e tropa.

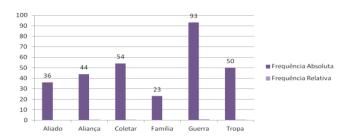


Figura 6: Frequência dos conceitos do primeiro resultado.

### 4.3 3º Resultado – Conceitos da Ferramenta Sobek aplicados na Text Mining Suite

O terceiro resultado referente a etapa 5 foi originado pela etapa 4 demostrada na Fig. 2 da metodologia, onde a mineração de texto da ferramenta Sobek gerou os conceitos demostrados na Fig. 5, desses, 8 substantivos (aliança, estado, guerra, jogo, pessoas, poder, tempo e tropa) foram adotados para definir os conceitos para aplicar na ferramenta text mining Suite.

Estado, jogo, pessoas, poder e tempo são os conhecimentos novos descobertos. No entanto, dos conceitos novos relevantes extraídos, 3 já tinham sido estabelecidos a partir do conhecimento e experiência do jogo, são eles: aliança, guerra e tropa.

Os conceitos novos resultantes da mineração de textos com a ferramenta Sobek somados aos conceitos definidos anteriormente, que aparaceram com mais frequência nos textos, propiciaram a definição de 9 conceitos (aliança, coletar, estado, guerra, jogo, pessoas, poder, tempo e tropa) para ser aplicados à ferramenta text mining. Como resultado, foram geradas 36 regras associativas, apresentadas na Tabela 2.

#### 4.4 Considerações sobre os testes realizados

O primeiro resultado do teste com a ferramenta Text Mining Suite, tendo os conceitos definidos pelo especialista (indivíduo que conhece termos referente ao jogo), descobriu as regras associativas entre os conceitos relacionadas à estratégia de jogadores.

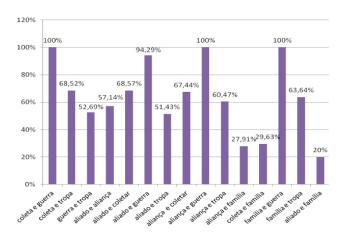
Nas Fig. 6 e Fig. 7 são retomados os dados do primeiro resultado, onde são apresentadas, respectivamente, a frequência com que os conceitos ocorrem no texto e as regras associativas com seus percentuais de confiança.

O resultado do segundo teste com a ferramenta Sobek foi concebido pelo relacionamento entre os conceitos mais frequentes da base de dados textual. Sendo que dos conceitos gerados pela ferramenta Sobek, 3 deles já eram conhecidos pelo especialista e estavam definidos como conceitos para o Text Mining Suite.

A Sobek apresentou 5 conceitos novos relevantes, que não eram de conhecimento do especialista. Isso contribuiu para a identificação de novas regras associativas e constatou-se que houve uma melhora na qualidade dos resultados gerados pelo Text Minig Suite, a partir da descoberta de novos conceitos pelo auxílio

**Tabela 2:** Associações entre conceitos no terceiro resultado

Confiança da Regra	Regras Associativas	Frequência
100%	aliança $ ightarrow$ guerra	43%
100%	coletar  ightarrow guerra	61%
96,43%	estado $ ightarrow$ guerra	27%
93,44%	coletar  o pessoas	57%
90,67%	$jogo \to pessoas$	68%
90,11%	$gerra \rightarrow pessoas$	82%
89,29%	estado $\rightarrow$ jogo	25%
89,29%	$estado \rightarrow pessoas$	25%
88,37%	aliança $ ightarrow$ jogo	38%
88,37%	aliança $ o$ pessoas	38%
80,33%	coletar  o jogo	49%
79,12%	gerra → jogo	72%
69,77%	aliança $ ightarrow$ coletar	30%
68,97%	$poder \rightarrow tropa$	20%
67,44%	aliança $ o$ tempo	29%
65,52%	$poder \rightarrow tempo$	19%
62,30%	coletar  o tempo	38%
62,30%	coletar  ightarrow tropa	38%
61,54%	guerra  o tempo	56%
60%	$jogo \rightarrow tempo$	45%
37,70%	coletar  o poder	23%
31,87%	guerra  o poder	29%
30,68%	$pessoas \rightarrow poder$	27%
59,09%	$pessoas \rightarrow tempo$	52%
57,14%	estado $\rightarrow$ tempo	16%
55, 81%	aliança $ ightarrow$ tropa	24%
51, 14%	$pessoas \rightarrow tropa$	45%
50,55%	guerra  o tropa	46%
48%	$jogo \to tropa$	36%
46,67%	tempo  o tropa	28%
41,86%	aliança $ ightarrow$ poder	18%
39,53%	aliança $ ightarrow$ estado	17%
39, 29%	estado  o tropa	11%
32%	$jogo \to poder$	24%
26,23%	$coleta \rightarrow estado$	16%
17,86%	$estado \rightarrow poder \\$	05%



**Figura 7:** Regras associativas do primeiro resultado.

da ferramenta Sobek, como mostrado na Tabela 2.

E, por fim, o terceiro teste resultou da definição dos conceitos da Sobek para a criação dos conceitos da

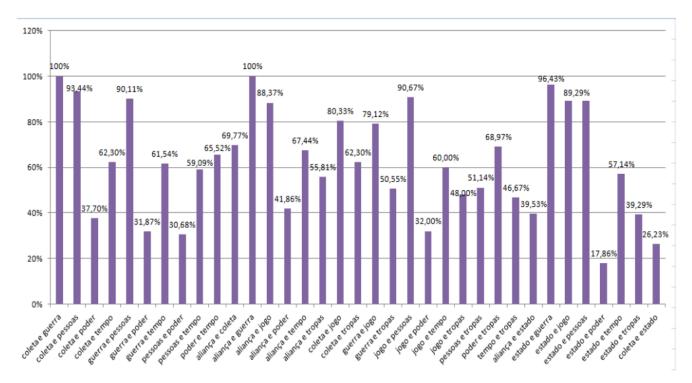


Figura 8: Regras associativas do terceiro resultado.

Text Mining Suite, a partir da seleção dos conceitos novos relevantes descobertos pela Sobek (estado, jogo, pessoas, poder e tempo) mais os conceitos criados pela compreensão do jogo (aliança, coletar, guerra e tropa). Na Fig. 8 são apresentadas as regras associativas com seus percentuais de confiança.

Pode se observar que o processo de mineração de texto semi-automático (Sobek + Text Mining Suite) produziu novas e melhores regras associativas, com maior grau de confiança (maior frequência). Com os conceitos descobertos foram identificados mais textos com os conceitos presentes e novas regras de associação.

Na Fig. 9 são exibidas as frequências dos conceitos nos textos. Salienta-se que os novos conceitos jogo, pessoas e tempo, aparecem com uma frequência significativa nos textos da coleção.

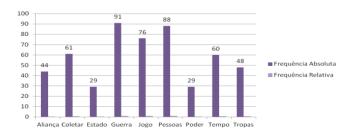


Figura 9: Frequência dos conceitos do primeiro resultado.

Algumas regras associativas apareceram com mais frequência na coleção, considerando o grau de confiança maior ou igual a 95% para os dois testes, como é mostrado na Fig. 10.

	Text Mining Suite	Sobek + Text Mining Suite
	- 100% dos textos que falam coleta também	- 100% dos textos que falam coleta
1	falam em guerra. Isso aconteceu em 54	também falam em guerra. Isso aconteceu
	textos (54% do total);	em 61 textos (61% do total);
	- 100% dos textos que falam aliança também	- 100% dos textos que falam aliança
2	falam em guerra. Isso aconteceu em 43	também falam em guerra. Isso aconteceu
	textos (43% do total);	em 43 textos 43% do total);
	- 100% dos textos que falam família também	- 96,43% dos textos que falam estado
3	falam em guerra. Isso aconteceu em 22	também falam em guerra. Isso aconteceu
	textos (22% do total);	em 27 textos 27% do total);

**Figura 10:** Comparativo das regras de associação.

Como interpretação dos resultados apresentados na Fig. 10, pode-se observar na regra de associação extraída pelo processo de descoberta de conhecimento: 100% dos textos que falam coleta também falam em querra. Isso aconteceu em 61 textos (61% do total), uma estratégia dos jogadores. A interpretação dessa regra é que guerras entre alianças ou entre estados tornam a atividade de coletar insegura, pois não há meio de proteger as tropas quando se coleta recursos, dessa forma, as tropas de um jogador ficam vulneráveis a ataques dos inimigos. A atividade de coleta é importante no jogo, pois nessa atividade, os recursos são recolhidos para alimentar e treinar soldados e realizar construções na

base do jogador, bem como ajudar os membros de sua aliança, se uma aliança fica impossibilitada de coletar, os jogadores não conseguem ter grande avanços no jogo.

A regra de associação: 100% dos textos que falam aliança também falam em guerra e a regra 96,53% dos textos que falam em estado também falam em guerra, indicam guerra entre alianças e guerra entre estados. As guerras entre alianças podem acontecer a qualquer momento do jogo, mas a guerra entre estados ocorre eventualmente (quando os desenvolvedores do jogo estabelecem). Isso justifica o percentual 43% e 27% do total dos textos das ocorrências destas regras.

#### 5 Conclusão

O presente trabalho apresentou o estudo do processo de descoberta de conhecimento aplicado em textos do chat do jogo MMO Mobile Strike, além disso, propôs uma metodologia para a aplicação das ferramentas de mineração de textos: Text Mining Suite e Sobek, que realizam as etapas da descoberta de conhecimento em textos, e, dessa forma, foi possível extrair conhecimento novo da coleção de textos.

A utilização da Text Mining Suite requer a compreensão do assunto na qual se pretende extrair conhecimento para determinar os conceitos presentes no texto, bem como, a Sobek encontra os relacionamentos entre os conceitos. Desta forma, a Sobek, além de encontrar os conceitos mais relevantes em um texto, pode auxiliar na concepção dos conceitos para a utilização da Text Mining Suite.

O conhecimento adquirido, através das regras extraídas pela ferramenta Text Mining Suite com o auxílio da ferramenta Sobek, podem ser aplicados como estratégias para os jogadores do jogo, como exemplo, não coletar quando alianças estão em guerra.

Essa pesquisa envolveu algumas limitações, como a coleta de textos foi realizada de forma manual, assim, seria interessante automatizar a coleta dos textos; também seria importante que diversos conceitos sob outras percepções pudessem ser criados, no sentido de ampliar a abrangência dos assuntos.

Como trabalhos futuros, propõe-se a extração de novos conhecimentos com enfoque diferentes, como comportamentos ou tendências de jogadores, visto que as conversas dos jogadores não são o tempo todo sobre o jogo em si.

#### Referências

- Almeida, C. E., Correia, R. C., Eler, D. M., Olivete-Jr, C., Garci, R. E., Scabora, L. C. and Spadon, G. (2017). Prediction of winners in moba games, 12th Iberian Conference on Information Systems and Technologies (CISTI), Ieee, pp. 1–6. https://doi.org/10.23919/CISTI.2017.7975774.
- Ambrozio, A. P. L. and Morais, E. A. M. (2007). Mineração de textos, *Technical Report 005/07*, Universidade Federal de Goiás. Disponível

- em http://ww2.inf.ufg.br/sites/default/files/
  uploads/relatorios-tecnicos/RT-INF\_005-07.pdf.
- Barbat, M. M., Adamatti, D. F. and Werhli, A. (2014). Desenvolvimento de um jogo sério para modelagem, operação e manutenção de plantas industriais, Revista Junior de Iniciação Científica em Ciências Exatas e Engenharia 7: 1-6. Disponível em http://www.icceeg.c3.furg.br/index.php?Itemid=837&option=bloco\_texto&id\_site\_componente=1241.
- Carvalho, A., Faceli, K., Lorena, A. and Gama, J. (2011). Inteligência Artificial—uma abordagem de aprendizado de máquina, Rio de Janeiro: LTC.
- Corrêa, A. C. G. et al. (2003). Recuperação de documentos baseados em informação semântica no ambiente ammo., Master's thesis, Universidade Federal de São Carlos, Brasil. Disponível em https://repositorio.ufscar.br/handle/ufscar/522.
- Epstein, D. and Reategui, E. B. (2015). Uso de mineração de textos no apoio à compreensão textual, RE-NOTE: revista novas tecnologias na educação [recurso eletrônico]. Porto Alegre, RS. https://doi.org/10.22456/1679-1916.57647.
- Fleury, A., Nakano, D. and Cordeiro, J. (2014). Mapeamento da indústria brasileira e global de jogos digitais, GEDIGames/USP, São Paulo, Brasil.
- Klemann, M., Reategui, E. and Lorenzatti, A. (2009). O emprego da ferramenta de mineração de textos sobek como apoio à produção textual, Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação-SBIE), Vol. 1. Disponível em https://www.br-ie.org/pub/index.php/sbie/article/view/1154.
- Loh, S. (2001). Abordagem baseada em conceitos para descoberta de conhecimento em textos, PhD thesis, Universidade Federal do Rio Grande do Sul. Instituto de Informática. Programa de Pós-Graduação em Computação. Disponível em https://www.lume.ufrgs.br/handle/10183/1849.
- Loh, S., Wives, L. K. and de Oliveira, J. P. M. (2000). Concept-based knowledge discovery in texts extracted from the web, ACM SIGKDD Explorations Newsletter 2(1): 29-39. https://dl.acm.org/doi/abs/10.1145/360402.360414.
- Martins, Y., Azevedo, B. F. T. and Bastos, H. P. P. (2010). Identificação de palavras compostas como auxílio à mineração de textos: Desenvolvimento do compostas\_v1, VI Circuito de Tecnologia de Informação CITI. Campos dos Goytacazes,RJ.
- Murnion, S., Buchanan, W. J., Smales, A. and Russell, G. (2018). Machine learning and semantic analysis of in-game chat for cyberbullying, *Computers & Security* **76**: 197–213. https://doi.org/10.1016/j.cose. 2018.02.016.
- Musabirov, I., Okopny, P. and Bulygin, D. (2015). Analyzing chat logs in online games for tutorial improvement, *Proceedings of the 2015 Annual*

- Symposium on Computer-Human Interaction in Play, ACM, pp. 661-666. https://dl.acm.org/doi/10.1145/2793107.2810335.
- Novak, J. (2010). *Desenvolvimento de games*, Vol. 1, Cengage Learning, São Paulo, Brasil.
- Rezende, S. O. (2003). Sistemas inteligentes: fundamentos e aplicações, Editora Manole Ltda, Barueri, São Paulo, Brasil.
- Schiessl, J. M. (2007). Descoberta de conhecimento em texto aplicada a um sistema de atendimento ao consumidor, Master's thesis, Universidade de Brasília, Faculdade de Economia, Administração, Contabilidade e Ciência da Informação e Documentação, Departamento de Ciência da Informação e Documentação. Disponível em https://repositorio.unb.br/handle/10482/1414.
- Soares, A. (2010). A mineração de texto na análise de contas públicas municipais, Master's thesis, State University of Ceará, Fortaleza, Brasil. Disponível em http://bdtd.ibict.br/vufind/Record/UECE-0\_a8203cec413e7c6f863a7490214063e3.
- Sowa, J. F. et al. (2000). Knowledge representation: Logical, Philosophical, and Computational Foundations, Brooks/Cole, Pacific Grove. Disponível em http://www. jfsowa.com/krbook/.
- Sperotto, F. A. (2013). Um modelo de suporte à comunicação de agentes para tratamento de informação imprecisa baseada em sinônimos, Master's thesis, Universidade Federal do Rio Grande, Brasil. Disponível em <a href="http://repositorio.furg.br/handle/1/6575">http://repositorio.furg.br/handle/1/6575</a>.
- Thompson, J. J., Leung, B. H., Blair, M. R. and Taboada, M. (2017). Sentiment analysis of player chat messaging in the video game starcraft 2: Extending a lexicon-based model, *Knowledge-Based Systems* 137: 149–162. https://doi.org/10.1016/j.knosys.2017.09.022.
- Wang, X. and Goh, D. H.-L. (2020). Components of game experience: An automatic text analysis of online reviews, *Entertainment Computing* **33**: 100338. https://doi.org/10.1016/j.entcom.2019.100338.
- Zambenedetti, C. (2002). Extração de informação sobre bases de dados textuais, Master's thesis, Universidade Federal do Rio Grande do Sul. Instituto de Informática. Programa de Pós-Graduação em Computação. Disponível em https://www.lume.ufrgs.br/handle/10183/1628.