

## Protocolo para análise de ferramentas de mineração de opiniões em redes sociais

### RESUMO

A mineração de opinião também conhecida como análise de sentimentos é uma área em constante evolução responsável por analisar opiniões sobre uma determinada entidade, serviço ou produto. Este trabalho propõe-se a criar um protocolo para comparar diferentes ferramentas que realizam a mineração de opinião nas redes sociais. Com base na literatura de apoio foi concebido o protocolo comparativo (constituído por sessenta questões) considerando critérios de recursos, usabilidade e características inerentes à tarefa de mineração de opiniões. Na sequência foram estudadas, testadas e selecionadas seis ferramentas que foram comparadas pelos critérios do protocolo proposto. Como resultado foi possível identificar que a ferramenta Seekr destacou-se perante as outras, principalmente por disponibilizar a análise em diversas redes sociais em tempo real e por permitir que a classificação da opinião seja realizada com o apoio de um analista. Por fim este trabalho julga de extrema importância que o estudo sobre a mineração de opinião tenha continuidade, uma vez que este assunto se configura como uma potencial área de trabalho aos profissionais de ciência da informação, ciência da computação e outras atividades relacionadas.

**PALAVRAS-CHAVE:** Mineração de dados. Recuperação de informação. Base de dados online.

**Jéssica Bonafini Berlim**

[jess.berlim@hotmail.com](mailto:jess.berlim@hotmail.com)

Universidade Federal do Paraná, Curitiba, Paraná, Brasil.

**Denise Fukumi Tsunoda**

[dtsunoda@ufpr.br](mailto:dtsunoda@ufpr.br)

Universidade Federal do Paraná, Curitiba, Paraná, Brasil.

## INTRODUÇÃO

Uma opinião pode ser a causadora de grandes constrangimentos, de extremo contentamento, de profunda influência ou até mesmo de uma grande descoberta. Essa mesma opinião pode gerar milhares e milhares de outras opiniões, sejam elas complementares ou opostas, compostas com argumentos para tal concordância ou oposição.

Uma opinião pode ensejar raiva, felicidade, contentamento, descontentamento, indiferença, tristeza dentre outros sentimentos. E com a ubiquidade das redes sociais, blogs, fóruns etc. muitas opiniões estão livremente disponíveis para análises.

Diante desse panorama, alinhado a importância de se compreender o que o consumidor, cliente, usuário está dizendo para conquistá-lo surgem as diversas tentativas de exploração de técnicas e ferramentas para análise das redes sociais. Assim, na busca por novas estratégias, para o auxílio na tomada de decisões e para se alcançar vantagem competitiva surge uma nova possibilidade: saber realmente, de maneira probabilística e comprobatória o anseio das pessoas. Essa possibilidade pode ser traduzida como o método “Mineração de Opinião” (MO), também aceito por diversos autores (apesar de não ser consenso) como “Análise de Sentimentos” (AS) e pode ser realizada de forma automática por meio de ferramentas disponíveis no mercado. Nesse sentido Pang e Lee (2008) afirmam que “O que as outras pessoas pensam sempre foi uma parte importante das informações para a maioria das pessoas durante o processo de tomada de decisão.”.

A mineração de opinião possibilita que a opinião de pessoas (e aqui está se falando de ultrapassar as barreiras geográficas por meio da Internet) no mundo todo possa ser avaliada e utilizada para tomada de decisões. Por meio deste método é possível descobrir qual é o sentimento das pessoas em relação a um candidato às eleições, como as pessoas repercutem o fato de um produto novo ser lançado no mercado, como opinam em relação a um serviço dentre outros. A mineração de opinião busca traçar um padrão na opinião das pessoas, por meio da análise textual, e identificá-lo como positivo, negativo ou neutro. A partir disso, coloca à disposição dos interessados, relatórios sobre determinado assunto, produto ou serviço.

O alto fator de influência de uma opinião pode ser determinante para que sejam estudados meios de transformar essas opiniões em estratégia e vantagem competitiva. Existem diversas ferramentas disponíveis para tal tarefa, no entanto é preciso conhecer as especificidades de cada uma, para que seja escolhida a mais adequada a uma dada situação. A tarefa de análise de opiniões é considerada complexa uma vez que uma opinião pode estar repleta de ambiguidade e ironia, e outros elementos que dificultam a análise destas informações. Daí surge a importância das ferramentas que viabilizam (ou pretendem viabilizar) a análise de opiniões de forma automática (ou mais automatizada possível) aliadas ao conhecimento do analista humano.

Segundo Liu (2012), uma das principais razões para a falta de estudo sobre opiniões é o fato de que havia pouco texto opinativo disponível antes da world wide web. Nos últimos anos, os estudos na área da análise de sentimentos

aumentaram justamente com o advento da internet e principalmente das redes sociais que são grandes polos de concentração das opiniões.

Silva et al, (2012, p.2) define os três níveis de granularidade da análise de sentimentos:

[...]pode ser realizada em 3 níveis de granularidade: (1) no nível de documento, observando o sentimento global expresso no texto; (2) no nível de sentença, classificando a polaridade de cada sentença no texto; e (3) no nível de característica, analisando a polaridade das opiniões sobre características/atributos dos objetos. (SILVA et al, 2012, p.2).

Segundo Ferreira (2010), “A classificação de opiniões na análise de sentimento é um tópico muito pesquisado, e está relacionado com a categorização de documentos ou sentenças, os quais expressam opiniões, como opinião positiva ou negativa.”. Esta é a etapa que define a pesquisa, pois aponta se a polaridade de uma determinada entidade ou um texto como um todo é positiva, negativa ou neutra.

Assim, este artigo apresenta um protocolo para a avaliação das ferramentas de mineração de opinião. Na sequência apresenta o resultado de um comparativo entre algumas ferramentas, evidenciando-se qualidades e fragilidades das mesmas identificadas pela aplicação do protocolo apresentado.

## METODOLOGIA

O protocolo foi criado a partir de pesquisas em referências bibliográficas, no trabalho de Análise Comparativa entre as Ferramentas OLAP de Sá et al (2012) e nos relatórios de análises e comparação de ferramentas disponibilizados pelo Centro de Recursos Computacionais (CERCOMP, 2016).

Para validação do protocolo proposto neste estudo foi concebida uma base de dados, por meio da utilização da ferramenta *Seekr*, a partir de opiniões coletadas com o termo “*Game of Thrones*” nas redes sociais Twitter e Facebook. No site da Wiki de *Game of Thrones*, são disponibilizadas informações de que a série de televisão americana foi criada por David Benioff e D. B. Weiss para a HBO. O seriado foi escolhido devido a sua grande popularidade em diversos países, incluindo o Brasil. Na época da coleta, a HBO estava transmitindo a 5ª temporada da série e por este motivo os episódios estavam sendo amplamente comentados e estavam divididos as opiniões dos expectadores. A busca pelo termo na ferramenta *Seekr* se deu um dia após o 6º episódio da 5ª temporada, “*Unbowed, Unbroken, Unbent*”, ser transmitido no Brasil.

Os resultados da pesquisa apontam qual o sentimento das pessoas em relação à série, mas não podem ser generalizados, uma vez que milhões de pessoas acompanham tal seriado e apenas 400 opiniões (Twitter e Facebook) foram coletadas para compor este estudo. Algumas ferramentas, nas versões utilizadas, não permitiram a utilização da base de dados concebida. Nestes casos, as bases foram criadas diretamente nas ferramentas, conforme explicado nas análises.

A escolha pelas ferramentas a serem submetidas ao protocolo se deu pelos seguintes critérios: possibilidade de busca de termos em redes sociais, seja pela

base criada no estudo ou pela própria ferramenta de busca inerente à solução; métodos diversificados na etapa de sumarização (visualização dos relatórios); possibilidade de identificação das etapas, níveis ou conceitos da análise de sentimentos. A ferramenta Social Mention foi utilizada para o pré-teste e ajustes das questões, por exemplo com a inclusão de questão sobre o retorno de suporte técnico, do protocolo. Finalmente, estas ferramentas estão listadas no Quadro 1.

QUADRO 1 – Ferramentas analisadas.

ID	Nome da Ferramenta	Versão da Ferramenta	Meio de Obtenção de Dados sobre a Ferramenta
1	Semantria	2013 x64 4.4.307	Aston, Liddle e Hu (2014), Lawrence (2013) e site do fornecedor
2	Text2data	3.5.1.360	Fórum Quora e site do fornecedor
3	Seekr		Fórum Quora e site do fornecedor
4	MeaningCloud	2.0.0.0	Fórum Quora e site do fornecedor
5	Umigon		Levallois (2013) e site do fornecedor
6	Sentiment140		Rosenthal <i>et al</i> (2015) e site do fornecedor

Fonte: Autoria própria, 2015.

Outras ferramentas também foram avaliadas, mas descartadas, conforme justificativas apresentadas no Quadro 2.

Quadro 2 – Ferramentas Estudadas E Descartadas Da Análise Pelo Protocolo

ID	Nome da Ferramenta	Meio de Obtenção de Dados sobre a Ferramenta
1	Attensity	Não foi possível testar. A versão demo foi requisitada, mas não houve resposta.
2	Clarabridge	Não foi possível testar. A versão demo foi requisitada, mas não houve resposta do fornecedor.
3	Pattern	Não foi testada pois, exige conhecimento em linguagem Phyton 2.5
4	TheySay	Testada mas não avaliada. Possui versão demo, mas não é possível conseguir resultados expressivos no teste para comparação com as outras ferramentas.
5	NLTK text classification	Testada mas não avaliada. Possui versão demo mas muitos recursos utilizados na comparação (pelo protocolo) não estão disponíveis nesta versão.
6	Tweetstats	Não foi possível testar, a página ficou carregando e não funcionou em vários testes.

Das ferramentas analisadas, a Semantria é uma solução criada pela companhia Lexalytics, conhecida pelo fornecimento de softwares de análise de textos. Conforme descrito no site da Lexalytics (SEMANTRIA, 2016) a empresa criou o primeiro motor comercial de análise de sentimentos. Segundo Lawrence (2014, p.2, tradução nossa), “Semantria é uma ferramenta de análise de sentimentos que pode ser aplicada a qualquer tipo de documentos, sentenças ou frases.”. Este afirma ainda que “[...]o software do sentimento, oferece uma aplicação baseada em um plugin do Excel, que possibilita a análise nas planilhas

do Excel e classifica os sentimentos como positivos, negativos e neutros.”. (LAWRENCE, 2014, p.5, tradução nossa). Por último o autor afirma que desde o lançamento da ferramenta em 2011, empresas e pesquisadores estão usando Semantria para realizar a Análise de sentimentos. (LAWRENCE, 2014).

Conforme descrito no site da Text2data (2016) a empresa criadora da ferramenta é uma *start-up* localizada em Londres que oferece serviços de análise de texto. A equipe tem mais de 10anos de experiência em engenharia de software e análise de dados, com foco em Business Intelligence e Pesquisa de Mercado.

Conforme descrito no site da Seekr (2016) a ferramenta nasceu em 2011 para atender a necessidade do mercado em acompanhar tudo o que é falado sobre as marcas, seus produtos e concorrentes dentro das mídias sociais. Em função desta necessidade, lançou o software de monitoramento, Seekr Monitor. A ferramenta é responsável por realizar análises nas redes sociais e se propõe a analisar o sentimento das opiniões coletadas.

Segundo o site Meaning Cloud (2016) a ferramenta é a maneira mais fácil, mais poderosa e mais acessível para extrair o significado de todo o tipo de conteúdo não estruturado. O site informa ainda que o usuário pode extrair automaticamente o significado de tweets, mensagens sociais, opiniões etc. Quanto à análise de sentimentos o site menciona que a automação da análise de sentimentos do Meaning Cloud permite processar dados que, devido ao seu volume, variedade e velocidade não podem ser tratados de forma eficiente pelos recursos humanos, pois consideram impossível extrair um valor pleno de interações utilizando tratamento exclusivamente manual.

Segundo Levallois (2013) a ferramenta Umigon foi desenvolvida em dezembro de 2012 com intuito de ser uma aplicação web que presta o serviço de detecção de sentimento, conhecido como mineração de opinião ou análise de sentimentos, em tweets. A Umigon foi concebida para ser rápida e escalável. Fornece indicações de traços semânticos adicionais presentes nos tweets, tais como indicações de tempo ou marcadores de subjetividade. É uma ferramenta on-line e sua utilização é relativamente simples quando comparada às demais ferramentas analisadas.

Segundo o site da Sentiment140 (2016) a ferramenta foi criada por três estudantes de pós-graduação em Ciência da Computação da Universidade de Stanford: Alec Vão, Richa Bhayani e Lei Huang. Foi concebida com o intuito de permitir descobertas de sentimentos relacionados à marca, produto ou tópico no Twitter. A abordagem da Sentiment 140 é considerada diferenciada de outros sites de análise de sentimento, uma vez que utiliza classificadores construídos a partir de algoritmos de aprendizado de máquina. Destacando-se em relação a sites que usam uma abordagem mais simples baseadas em palavras-chave.

## DESCRIÇÃO DO ESTUDO

Para efetivar as comparações das ferramentas de mineração de opiniões, foi criado um protocolo comparativo que contemplou questões relativas aos recursos, usabilidade e características inerentes à tarefa de mineração de opiniões:

a) recursos: foram avaliados os recursos disponibilizados por cada ferramenta, a fim de orientar os usuários em relação as dificuldades e facilidades que serão encontradas na utilização da ferramenta;

b) usabilidade: seção do protocolo que visa mostrar até que ponto a interface de uma ferramenta facilita ou dificulta seu entendimento, levando em conta pontos cruciais que podem fazer a diferença desde a instalação até a análise dos resultados;

c) características: neste quesito o protocolo visa investigar as características da ferramenta em relação a mineração de opinião, ou seja, se a ferramenta executa todas as atividades, etapas e níveis da mineração de opinião. Busca ainda demonstrar funcionalidades que ultrapassam as expectativas e destacar pontos que possam interferir de alguma maneira a análise.

Para assegurar a consistência na interpretação dos critérios do protocolo proposto foram criados glossários explicativos. O primeiro glossário (Quadro 3) apresenta o significado das questões elaboradas quanto as características gerais e recursos de suporte técnico e documentações disponíveis nas ferramentas.

Quadro 3 – Explicações dos critérios utilizados quanto às características gerais e recursos de suporte e documentação da ferramenta

<b>Recursos</b>	<b>Significado</b>
<b>Gratuita?</b>	A ferramenta está disponível para download gratuitamente.
<b>Paga?</b>	É necessário pagar um determinado valor para baixar e utilizar a ferramenta.
<b>Versão Demo?</b>	Existe uma versão de demonstração gratuita com recursos limitados.
<b>Versão Trial?</b>	Existe uma versão para teste, liberada para uso por um limite de tempo pré-determinado.
<b>Quanto tempo para usar a versão Trial?</b>	Especifica o tempo para utilização da versão Trial.
<b>Online?</b>	A ferramenta está disponível para uso on-line e não necessita que seja efetuado download.
<b>É necessário download?</b>	A ferramenta deve necessariamente ser instalada no computador e por isso é disponibilizada para download.
<b>Fácil instalação?</b>	A ferramenta executa a instalação de maneira intuitiva e sequencial, sem necessidade de um conhecimento aprofundado.
<b>No momento da instalação outras ferramentas plugins ou programas são disponibilizadas para download?</b>	A instalação da ferramenta é interrompida em alguns momentos para questionar se o usuário deseja instalar junto com a ferramenta outros dispositivos.
<b>A instalação é rápida?</b>	Para instalar a ferramenta são necessários apenas alguns minutos.
<b>É possível escolher um idioma?</b>	A ferramenta está disponível em vários idiomas e possibilita a seleção de um deles para uso.
<b>É possível utilizar a ferramenta em português?</b>	A ferramenta é disponibilizada no idioma português.
<b>Possui suporte no site?</b>	A ferramenta disponibiliza suporte on-line (chat, e-mail) no site para sanar as dúvidas.

<b>Possui link de suporte dentro da ferramenta?</b>	A ferramenta disponibiliza um link direto o serviço de suporte.
<b>O suporte executa efetivamente seu papel?</b>	Quando questionados, os responsáveis pelo suporte atendem prontamente as dúvidas do usuário.
<b>Disponibiliza tutorial?</b>	A ferramenta disponibiliza em seu site um tutorial com o passo-a-passo para sua utilização.

Fonte: Autoria própria, 2015.

O Quadro 4 apresenta o significado das questões levantadas quanto a usabilidade, ou seja, serão abordadas questões que permitam ao usuário saber se as ferramentas oferecem conforto e fácil entendimento durante o uso.

Quadro 4 - Explicações dos critérios utilizados quanto à usabilidade da ferramenta

<b>Usabilidade</b>	<b>Significado</b>
<b>Possui design intuitivo?</b>	O design da ferramenta permite que se entenda cada função de maneira lógica, sem precisar consultar tutoriais ou solicitar auxílio do suporte.
<b>É fácil de usar?</b>	É fácil para um usuário iniciante identificar suas funcionalidades, onde encontra-las e como executá-las.
<b>A ferramenta é atrativa?</b>	A interface da ferramenta é amigável e atrativa.
<b>É ágil?</b>	A ferramenta funciona com rapidez, sem travar ou ter que reiniciá-la.
<b>É possível executar a ferramenta?</b>	A ferramenta funciona após instalação, sem mensagens de erro.

Fonte: Autoria própria, 2015.

O Quadro 5 apresenta a explicação das questões abordadas quanto as características da ferramenta, essas perguntas auxiliarão o usuário a entender o funcionamento da ferramenta durante as etapas da mineração de opinião.

Quadro 5 - Explicações Dos Critérios Utilizados Quanto às características Técnicas Inerentes Ao Processo De Mineração De Opiniões Disponíveis Na Ferramenta

<b>Características</b>	<b>Significado</b>
<b>A ferramenta permite o nivelamento prévio da mineração?</b>	É possível selecionar um tipo de nível antes de começar a fazer a mineração de opinião. Ou seja, a ferramenta permite que se defina se a mineração será feita à nível de documento, sentença ou característica.
<b>A ferramenta explicita os níveis da mineração a serem executados?</b>	A ferramenta demonstra de maneira explícita em que níveis foram executadas a mineração de opinião.
<b>A mineração é feita a nível de documento?</b>	A ferramenta possibilita que a mineração de opinião seja realizada à nível de documento, ou seja, a opinião é minerada como um todo.
<b>A mineração é feita a nível de sentença?</b>	A ferramenta possibilita que a mineração de opinião seja realizada à nível de sentença, ou seja, as frases opinativas são analisadas separadamente.
<b>A mineração é feita a nível de</b>	A ferramenta possibilita que a mineração de



<b>característica?</b>	opinião seja realizada à nível de característica, ou seja, cada aspecto de uma sentença pode ser analisado separadamente.
<b>A ferramenta permite realizar a etapa de subjetividade?</b>	A ferramenta permite a realização da etapa de subjetividade, onde são filtradas apenas as sentenças que se configuram como opiniões e são descartadas as que não são opiniões.
<b>A ferramenta permite realizar a etapa de extração?</b>	A ferramenta permite a realização da etapa de Extração, responsável pela extração de características de cada objeto contido nas sentenças.
<b>A ferramenta permite realizar a etapa de classificação?</b>	A ferramenta permite a realização da etapa de Classificação, que define a polaridade das opiniões, ou seja, se elas são positivas, negativas, ou neutras.
<b>A ferramenta permite realizar a etapa de sumarização?</b>	A ferramenta permite a realização da etapa de Sumarização, fase responsável por associar e expor os resultados da mineração de opinião.
<b>As etapas são realizadas sequencialmente?</b>	As etapas, subjetividade, extração, classificação e sumarização são executadas pela ferramenta de maneira sequencial.
<b>É possível realizar uma etapa com apoio de outras ferramentas?</b>	A ferramenta não disponibiliza os recursos para executar todas as etapas, mas permite que seja utilizada outra ferramenta em conjunto para que a mineração de opinião ocorra por completo, não sendo esse recurso de uso obrigatório.
<b>É necessário que se realizem etapas com apoio de outras ferramentas?</b>	A ferramenta necessariamente utiliza outra ferramenta em conjunto para que a mineração de opinião ocorra por completo.
<b>A ferramenta se restringe a minerar um campo específico?</b>	Dentro da ferramenta o conteúdo de pesquisa é previamente definido não sendo possível realizar a pesquisa em um determinado assunto de interesse. Ex. Ferramenta que minera opiniões sobre política.
<b>A ferramenta minera opiniões relacionadas a campos diferentes?</b>	A ferramenta permite que sejam feitas pesquisas de opiniões sobre assuntos diversos, não se restringe à um determinado campo.
<b>A ferramenta possibilita a mineração em diferentes plataformas?</b>	A ferramenta permite que a mineração de opinião seja realizada em diferentes plataformas. Ex. Facebook, blogs, Twitter, fóruns.
<b>A ferramenta possibilita a mineração em uma única plataforma?</b>	A ferramenta restringe a mineração de opinião em apenas uma plataforma. Ex. A ferramenta só minera opiniões do Twitter.
<b>A ferramenta realiza a mineração de maneira automática?</b>	Todas as etapas da mineração de opinião são feitas de maneira automática sem que seja necessária a intervenção de um usuário.



<b>A ferramenta realiza a mineração de forma semiautomática?</b>	Algumas etapas da mineração de opinião são realizadas manualmente com o auxílio de um usuário, enquanto outras são realizadas automaticamente.
<b>A etapa de subjetividade é realizada de maneira automática?</b>	A etapa de subjetividade é realizada de maneira automática sem que seja necessária a intervenção de um usuário.
<b>A etapa de extração é realizada de maneira automática?</b>	A etapa de extração é realizada de maneira automática sem que seja necessária a intervenção de um usuário.

Fonte: Autoria própria, 2015.

O Quadro 6 apresenta as características analisadas nas ferramentas com enfoque técnico da mineração de opiniões. Por exemplo, aspectos de pré-processamento de dados tais como extração são avaliados. Ainda, são avaliados os formatos de arquivos nos quais os resultados podem ser exportados.

Quadro 6 – Explicações Dos Critérios Utilizados Quanto Às Funcionalidades Inerentes Ao Processo De Mineração De Opiniões Disponíveis Na Ferramenta

<b>Características</b>	<b>Significado</b>
<b>A etapa de classificação é realizada de maneira automática?</b>	A etapa de classificação é realizada de maneira automática sem que seja necessária a intervenção de um usuário.
<b>A etapa de sumarização é realizada de maneira automática?</b>	A etapa de sumarização é realizada de maneira automática sem que seja necessária a intervenção de um usuário.
<b>A etapa de subjetividade é realizada manualmente?</b>	A etapa de subjetividade é realizada manualmente pelo usuário da ferramenta.
<b>A etapa de extração é realizada manualmente?</b>	A etapa de extração é realizada manualmente pelo usuário da ferramenta.
<b>A etapa de classificação é realizada manualmente?</b>	A etapa de classificação é realizada manualmente pelo usuário da ferramenta.
<b>A etapa de sumarização é realizada manualmente?</b>	A etapa de sumarização é realizada manualmente pelo usuário da ferramenta.
<b>A ferramenta recupera informações relevantes para se chegar à uma conclusão?</b>	O relatório disponibilizado pela ferramenta na última fase da mineração é conclusivo e traz informações que podem ser decisivas à tomada de decisão.
<b>O número de opiniões recuperadas é passível de análise?</b>	A quantidade de opiniões recuperadas pela ferramenta é suficiente para formar conclusões.
<b>A etapa de sumarização gera um relatório que pode ser exportado?</b>	O relatório gerado na última etapa da mineração de opinião está disponível para exportação.
<b>O relatório pode ser exportado em pdf?</b>	O relatório gerado na última etapa da mineração de opinião está disponível para exportação em pdf.
<b>O relatório pode ser exportado em MS-Word?</b>	O relatório gerado na última etapa da mineração de opinião está disponível para exportação em MS-Word.
<b>O relatório pode ser exportado em MS-Excel?</b>	O relatório gerado na última etapa da mineração de opinião está disponível para exportação em MS-Excel.
<b>O relatório pode ser exportado em outros programas?</b>	O relatório gerado na última etapa da mineração de opinião está disponível para exportação em outros programas.
<b>É possível conectar a ferramenta às contas pessoais nas redes</b>	A ferramenta permite a sincronização das contas das redes sociais do usuário para fins de análise.

<b>sociais?</b>	
<b>É possível fazer a análise com mais de um termo referente ao mesmo assunto?</b>	A ferramenta disponibiliza a busca por termos que abrangem um campo específico. Exemplo: Game of Thrones, GOT.
<b>É possível restringir a curva de tempo de busca de um termo?</b>	A ferramenta permite que se delimite uma data de início e fim para a busca de um termo.
<b>A ferramenta limita o número de buscas?</b>	A ferramenta restringe um número específico de buscas, ou seja, é delimitado pela ferramenta o número máximo de opiniões a serem recuperadas.
<b>É possível adicionar termo ignorado?</b>	É possível adicionar termos que possam confundir a ferramenta na hora da busca, evitando que sejam recuperadas informações não relevantes. Exemplo: A busca é sobre a maçã, é possível adicionar ao termo ignorado “do amor”, “do rosto”, “vermelha”.

Fonte: Autoria própria, 2015.

## RESULTADOS: APRESENTAÇÃO E DISCUSSÃO

Para o Semantria, o primeiro ponto positivo importante constatado na ferramenta é o fato da mesma funcionar dentro da planilha da Microsoft®, o MS-Excel®. A ferramenta mostrou-se bastante completa por realizar a análise em todos os níveis (documento, sentença e aspecto) e por detalhar todos os níveis e suas respectivas polaridades.

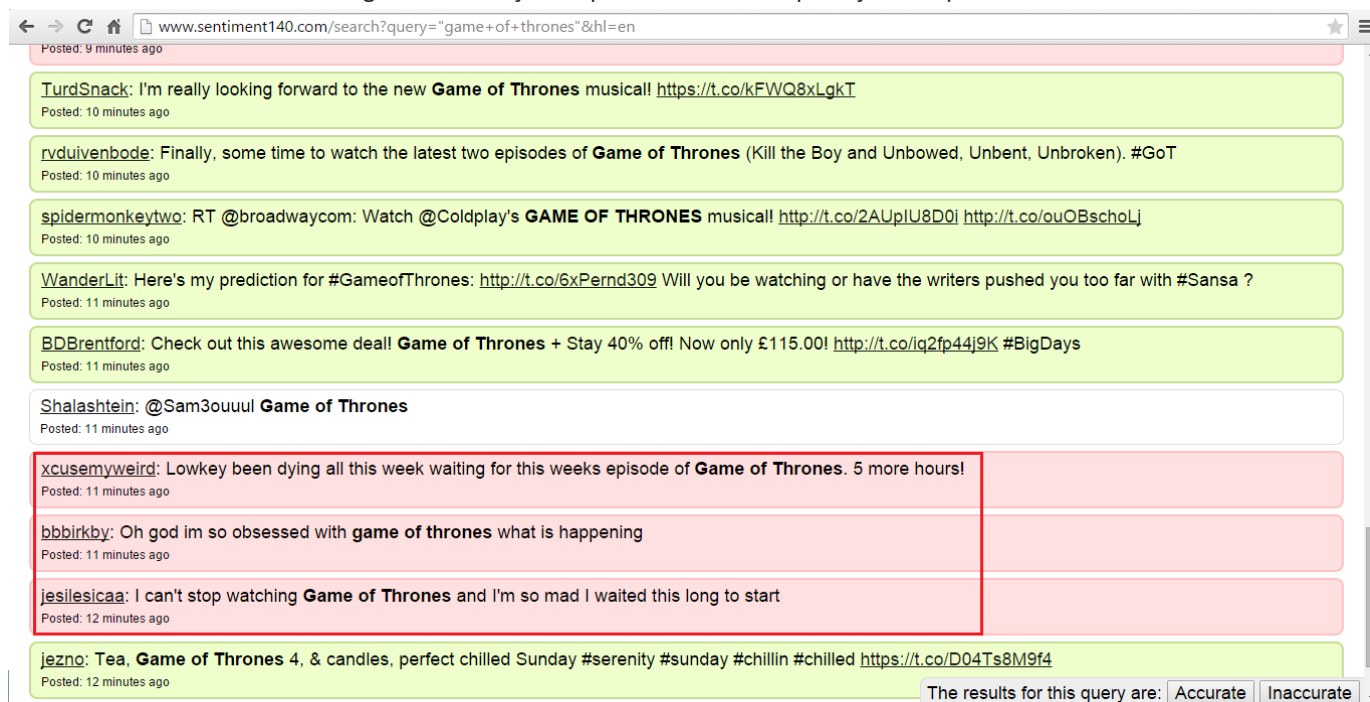
Apesar das bases para treino disponibilizadas pela Semantria não terem sido utilizadas, representa um ponto positivo pois não é preciso criar uma base própria para o estudo preliminar da mesma. Os tutoriais contidos no site são de fáceis compreensão e efetivos quando considerado o detalhamento dos tópicos abordados.

Outro ponto positivo é a visualização da informação na etapa de sumarização, em forma de relatórios e gráficos, permitindo ainda que o relatório seja exportado em vários tipos de arquivos. Além disso, a ferramenta realiza a análise de sentimentos no idioma português e outros sete idiomas.

Como fatores negativos podem ser destacados o fato da ferramenta ser paga e a versão gratuita ser disponibilizada apenas por 30 dias. Outro ponto observado foi o fato de que algumas ironias e gírias não serem detectadas quando a análise é realizada ao nível de documento.

Apenas para ilustração de alguns problemas de interpretação de ironias, gírias e outras particularidades, apresenta-se a Figura 1. As três frases em destaque utilizam termos que convencionalmente estão relacionadas às opiniões negativas: “*dying*” e “*obsessed*”. Na primeira frase destacada a pessoa afirma que está “morrendo” a semana toda esperando para ver o episódio de *Game of Thrones* e que faltam apenas 5 horas para começar – a frase é considerada como negativa quando na verdade é positiva, pois demonstra que a pessoa está ansiosa para ver o próximo episódio da série. Na segunda frase em destaque o usuário se pergunta – como pode estar tão obcecado por *Game of Thrones* [...] – também é considerada negativa, mas para a série é positiva, já que se subentende que a pessoa não consegue parar de assistir. Na última frase destacada a pessoa afirma que – não pode parar de assistir *Game of Thrones* [...] – mais uma vez a frase é considerada negativa, o que não é o caso.

Figura 1: Ilustração de problemas de interpretação de opiniões



Fonte: Sentiment140, 2015.

Em uma análise global a ferramenta se mostrou eficiente, principalmente por fazer uma análise em todos os níveis e por apresentar os resultados tabulados.

A ferramenta Text2Data analogamente a Semantria, trabalha de forma integrada à planilha da Microsoft®, o MS-Excel®. A ferramenta também disponibiliza bases para estudos, inclusive uma das bases com extrações do Twitter. As bases que acompanham a ferramenta não foram utilizadas neste estudo.

Toda a análise é detalhada, por isso é possível compreender que a análise é feita em todos os níveis e que todas as etapas da mineração de opinião são realizadas. Além disso, define a subjetividade dos documentos, o que foi um diferencial da ferramenta. Tutoriais estão disponíveis no site, porém o vídeo tutorial é um pouco confuso, pois não tem som e não mostra alguns passos essenciais para o início da mineração.

A etapa de sumarização é disponibilizada em forma de relatório, gráfica e ilustrativa. Permite que o relatório seja exportado em vários tipos de arquivos. Um dos pontos negativos é o fato da ferramenta, na versão que foi testada, exigir algumas versões antigas de sistemas operacionais.

A ferramenta *Seekr* permite a busca de termos em diferenciadas fontes na web, a exemplo das redes sociais, objeto de estudo desta pesquisa. Outro ponto positivo para a ferramenta é o fato de que as buscas podem ser realizadas em um período pré-definido. Isso pode ser útil, por exemplo, em épocas de eleições, pois a busca pode ser restrita ao período que antecede as eleições para se descobrir qual é o sentimento das pessoas em relação a determinado candidato. Outro ponto a ser registrado é que na base *Game of Thrones* não foi definido um período específico, mas os resultados estavam atualizados, uma vez que a

maioria dos tweets e Facebook posts recuperados abordavam o último episódio, exibido na véspera da coleta das opiniões. A permissão de 1000 ocorrências de busca é interessante para uma ferramenta apenas em versão de teste (período de 7 dias).

Um ponto negativo a ser apontado é que a polarização realizada de maneira automática na ferramenta apresenta problemas pois muitas frases não foram polarizadas de maneira correta. A ferramenta não conseguiu identificar ironias, abreviaturas e até algumas frases que não possuíam cujas opiniões não estavam explicitamente registradas.

O contato com a empresa *Seekr* é muito eficiente e o fato da ferramenta ser brasileira facilitou a interação. Logo que é realizado um login na ferramenta, um e-mail é recebido. Para tirar qualquer dúvida basta apenas responder o e-mail com a pergunta desejada. Houve um momento em que os monitoramentos não estavam aparecendo e foram enviadas telas apresentando a situação para o suporte que respondeu e esclareceu a situação rapidamente. O fato de buscar a base de dados on-line torna a ferramenta bastante versátil quando comparada às demais avaliadas.

De maneira geral a ferramenta *Meaning Cloud* mostrou-se efetiva e útil, alguns diferenciais tais como dividir a polaridade em cinco tipos (o que permite que a ferramenta saiba diferenciar uma frase muito positiva de uma frase apenas positiva) demonstram a tentativa da ferramenta em aumentar a precisão em sua análise. No entanto mais uma vez as ironias e gírias se mostraram um problema, pois muitas frases foram classificadas de maneira incorreta.

O fato de ter um tutorial detalhando todas as etapas a serem realizadas é um ponto positivo principalmente para usuários pouco familiarizados com o processo de mineração de opinião como um todo.

A ferramenta *Umigon* mostrou ser funcional quanto à coleta de dados uma vez que recupera os tweets relacionados à pesquisa e interpreta aqueles que são colocados na caixa de texto. No entanto deixa a desejar quando a análise é realizada a nível de documento já que não apresenta o motivo de alguns tweets não serem classificados.

É possível compreender também que a análise é realizada a nível de sentença. Além disso, o fato da ferramenta realizar as pesquisas no Twitter, demonstra que a ferramenta se preparou para interpretar “*hashtags*” e “*emoticons*”.

A leitura do artigo disponível no site é essencial para que se entenda como a ferramenta foi construída, um diferencial disponibilizado pela *Umigon*. Em contrapartida é preciso destacar a falta de informação em alguns quesitos, como escolha do idioma e relatórios para bases de dados que não são feitas pela ferramenta.

Como a ferramenta *Sentiment140* funciona como um site de buscas, concebida apenas para pesquisa on-line de tweets. Por isso não é preciso nada mais do que carregar o site para iniciar os procedimentos de análise. No entanto é importante deixar claro que a ferramenta só analisa tweets, e que os mesmos são recuperados pela própria ferramenta. Não é possível saber em que período as buscas estão sendo efetuadas e nem de que universo ou população as amostras estão sendo recuperadas. A ferramenta mostrou-se eficiente no que diz

respeito ao que se propõe, mas novamente trouxe problemas de interpretação quando surgem ironias e gírias.

## CONSIDERAÇÕES FINAIS

Pelas pesquisas realizadas foi possível identificar que as redes sociais deixaram de ser apenas um ambiente de relacionamento, para tornarem-se uma ferramenta de compartilhamento de informações e opiniões. Estas opiniões, por sua vez, estão sendo consideradas para decisões empresariais.

Assim, é possível afirmar em tempos em que a informação se mostra essencial e de valor inestimável para as organizações, minerar opiniões se torna um diferencial competitivo. Estudar as ferramentas que permitem a extração e os processos inerentes às tarefas de mineração de opiniões é essencial para facilitar as análises a serem realizadas pois a quantidade de dados aleatórios e disponíveis em profusão na rede comprova a necessidade de utilização de meios que ajudem a filtrar e analisar os dados pertinentes e relevantes. A pesquisa confirmou também a importância de um analista nesse processo, pois o mesmo pode interpretar os elementos da linguagem humana tais como ironias e gírias e “treinar” as ferramentas concebidas para este propósito.

Para facilitar a escolha das ferramentas, apresentou-se um protocolo para avaliação das ferramentas contemplando critérios quanto aos recursos, à usabilidade e as características inerentes ao processo de mineração de opiniões.

O protocolo proposto foi validado pela análise de seis ferramentas. Os resultados demonstraram que todas as ferramentas analisadas apresentam problemas significativos no que diz respeito à interpretação de gírias e ironias. Porém a *Seekr* foi considerada a melhor ferramenta se comparada as outras, devido aos seus variados recursos e por permitir que as classificações incorretas sejam alteradas (corrigidas) pelo usuário (analista) para que a ferramenta possa “acertar mais” em novas previsões e análises.

Dentro desse universo, foi possível perceber que a ferramenta utilizada para a realização da mineração de opinião é peça fundamental nesse processo, haja vista que cada etapa é repleta de complexidades específicas. Além disso é preciso ressaltar que existem inúmeras ferramentas a serem analisadas, como exemplo pode-se citar: as que analisam o sentimento das pessoas somente referente à comida; empresas que criaram ferramentas apenas para analisar opiniões sobre seu próprio produto ou serviço; outras apenas para pesquisa política, dentre outras. Este cenário demonstra o potencial de utilização do protocolo – inédito – proposto em novas versões das ferramentas analisadas ou ainda de outras ferramentas, representando grande contribuição para a área de análise de sentimentos.

## Analysis protocol for opinion mining tools in social web

### ABSTRACT

Opinion mining also known as sentiment analysis is an area in constant evolution, responsible for analyzing opinions on a given entity, service or product. This paper proposes the creation of a protocol to compare different tools that perform opinion mining on social networks. Exploratory work was performed in a few steps. Based on supporting literature it was designed a comparative protocol (made up of 60 questions), considering resources, usability and features related to opinion mining. Following, six tools were selected, studied, tested and subsequently compared under the developed protocol. The comparison and analysis allowed to learn that the tool Seekr stood out against the others, mainly by providing real time analysis on various social networks and allowing the opinion classification to be carried out with the support of an analyst. Finally, this work considers of utmost importance that the study on opinion mining continues since it has the potential be a work area of information science, computer science and others related professions.

**KEYWORDS:** Data mining. Information recovery. Online databases.

---

## REFERÊNCIAS

ASTON, Nathan; LIDDLE, Jacob; HU, Wei. Twitter sentiment in data streams with perceptron. **Journal of Computer and Communications**, 2014.

CERCOMP. Centro de Recursos Computacionais. Disponível em: <<https://www.cercomp.ufg.br/up/18/o/Jeremia.odt>>. Acesso em: 27 maio 2016.

FERREIRA, Emanuel de Barros Albuquerque. Análise de sentimento em redes sociais utilizando influência das palavras. 2010. 69 f. **Monografia** (Graduação) - Curso de Ciência da Computação, Centro de Informática, Universidade Federal de Pernambuco, Recife, 2010. Disponível em: <<http://www.cin.ufpe.br/~tg/2010-2/ebaf.pdf>>. Acesso em: 05 set. 2015.

LAWRENCE, Linus. **Reliability of sentiment mining tools**: a comparison of Semantria and Social Mention. 2014.

LEVALLOIS, Clement. Umigon: sentiment analysis for tweets based on lexicons and heuristics. In: **Proceedings of the International Workshop on Semantic Evaluation**, SemEval. 2013.

LIU, Bing. **Sentiment analysis and opinion mining: synthesis lectures on human language technologies**. [s.i.]: Morgan & Claypool Publishers, 2012. 168 p. Disponível em: <<http://www.cs.uic.edu/~liub/FBS/SentimentAnalysis-and-OpinionMining.pdf>>. Acesso em: 12 ago. 2015.

MEANING CLOUD. 2016. Disponível em: <<https://www.meaningcloud.com>>. Acesso em: 27 nov.2016.

PANG, Bo; LEE, Lillian. Opinion mining and sentiment analysis. In: **Foundations and trends in information retrieval**, v. 2, n. 1-2, p. 1-135, 2008.

ROSENTHAL, Sara et al. Semeval-2015 task 10: sentiment analysis in twitter. In: **Proceedings of the 9th International Workshop on Semantic Evaluation**, SemEval. 2015.

SÁ, Thays et al. Uma Análise Comparativa entre as Ferramentas OLAP como apoio a Soluções de BI nas Empresas. In: **V Congresso Tecnológico InfoBrasil**. Fortaleza, CE.2012.

SEEKR. 2016. Disponível em: <<http://seekr.com.br/>>. Acesso em: 27nov.2016.



SEMANTRIA. **Text analytics**, 2016. Disponível em: <<https://www.lexalytics.com/>>. Acesso em: 27 nov.2016.

SENTIMENT 140: discover the Twitter sentiment for a product or brand. 2016. Disponível em: <<http://www.sentiment140.com/>>. Acesso em: 27 nov.2016.

SILVA, Nelson Rocha; LIMA, Diego; BARROS, Flávia. SAPair: um processo de análise de sentimento no nível de característica. In: IV international workshop on web and text intelligence, 4., 2012, Curitiba. **Anais...** Curitiba: Anais, 2012. p. 1 - 10.

TEXT2DATA. 2016. Disponível em: <<http://text2data.org/>>. Acesso em: 27 nov.2016.

UMIGON: **sentiment analysis for tweets, and more**. 2016. Disponível em: <<https://github.com/seinecle/Umigon>>. Acesso em: 27 nov.2016.

**Recebido:** 28 mai. 2016.

**Aprovado:** 28 dez. 2016.

**DOI:** 10.3895/rts.v13n28.4002

**Como citar:** BERLIM, J. B.; TSUNODA, D. F. Protocolo para análise de ferramentas de mineração de opiniões em redes sociais. **R. Tecnol. Soc.**, Curitiba, v. 13, n. 28, p. 191-206, mai./ago. 2017. Disponível em: <<https://periodicos.utfpr.edu.br/rts/article/view/4002>>. Acesso em: XXX.

**Correspondência:**

Jéssica Bonafini Berlim.

Av. Bento Munhoz da Rocha Neto, 2289

Bairro: Jardim América

Paranaguá - PR

Telefone: (41)99758-5930

**Direito autorial:** Este artigo está licenciado sob os termos da Licença Creative Commons-Atribuição 4.0 Internacional.



## APÊNDICE - APLICAÇÃO DAS FERRAMENTAS AO PROTOCOLO COMPARATIVO

Recursos	Semantria	Seekr	Text2data	MeaningCloud	Umigon	Sentiment140
Gratuita?	Não	Não	Não	Não	Sim	Sim
Paga?	Sim	Sim	Sim	Sim	Não	Não
Versão Demo?	Sim	Não	Sim	Sim	Não	Não
Versão Trial?	Sim	Sim	Sim	Sim	Não	Não
Quanto tempo para usar a versão trial?	Não está informado	Após acabar o uso dos 1000 monitoramentos	30 dias	Não está informado	n/a	n/a
OnLine?	Não	Sim	Não	Não	Sim	Sim
É necessário download?	Sim	Não	Sim	Sim	Não	Não
Facil Instalação?	Sim	n/a	Não	Sim	n/a	n/a
No momento da instalação outras ferramentas plugins ou programas são disponibilizadas para download?	Não	n/a	Não	Não	n/a	n/a
A instalação é rápida?	Sim	n/a	Sim	Sim	n/a	n/a
É possível escolher um idioma?	Sim	Sim	Não	Não	Não	Sim
É possível utilizar a ferramenta em português?	Sim	Sim	Não	Não	Não está explícito	Não
Possui suporte no site?	Sim	Sim	Sim	Sim	Não	Não
Possui link de suporte dentro da ferramenta?	Sim	Sim	Sim	Sim	n/a	Não
O suporte executa efetivamente seu papel?	Sim	Sim	Sim	Sim	n/a	n/a
Disponibiliza tutorial?	Sim	Sim		Sim	Não	Não
Usabilidade	Semantria	Seekr	Text2data	MeaningCloud	Umigon	Sentiment140
Possui design intuitivo?	Sim	Sim	Sim	Sim	Sim	Sim

É fácil de usar?	Sim	Sim	Sim	Sim	Sim	Sim
A ferramenta é atrativa?	Sim	Sim	Sim	Sim	Não	Sim
É ágil?	Sim	Sim	Sim	Sim	Sim	Sim
É possível executar a ferramenta?	Sim	Sim	Apenas em sistemas operacionais de 32 bits	Sim	Sim	Sim
Características	Semantria	Seekr	Text2data	MeaningCloud	Umigon	Sentment140
A ferramenta permite o nivelamento prévio da mineração?	Não	Não	Não	Não	Não	Não
A ferramenta explícita os níveis da mineração a serem executados?	Sim	Não	Sim	Sim	Não	Não
A mineração é feita a nível de documento?	Sim	Sim	Sim	Não	Sim	Sim
A mineração é feita a nível de sentença?	Sim	Não	Sim	Sim	Sim	Sim
A mineração é feita a nível de característica?	Sim	Não	Sim	Sim	Não	Não
A ferramenta permite realizar a etapa de subjetividade?	Sim	Sim	Sim	Sim	Não está explícito.	Não está explícito.
A etapa de subjetividade está explícita?	Não	Não	Não	Sim	Não	Não
A ferramenta permite realizar a etapa de extração?	Sim	Sim	Sim	Sim	Sim	Sim
A etapa de extração está explícita?	Sim	Não	Sim	Sim	Não	Não
A ferramenta permite realizar a etapa de classificação?	Sim	Sim	Sim	Sim	Sim	Sim
A etapa de classificação está explícita?	Sim	Sim	Sim	Sim	Sim	Sim
A ferramenta permite realizar a etapa de sumarização?	Sim	Sim	Sim	Sim	Sim	Sim
A etapa de sumarização está explícita?	Sim	Sim	Sim	Sim	Sim	Sim
As etapas são realizadas sequencialmente?	Não está explícito	Não está explícito	Não está explícito	Sim	Não está explícito	Não está explícito
É possível realizar uma etapa com apoio de outras ferramentas?	Não	Sim, redes sociais	Não	Não	Sim, Twitter	Sim, Twitter
É necessário que se realizem etapas com apoio de outras ferramentas?	Não	Sim	Não	Não	Não	Sim
A ferramenta se restringe a minerar um campo específico?	Não	Não	Não	Não	Sim, apenas	Sim, apenas twets.

A ferramenta se restringe a minerar em uma plataforma específica?	Não	Não	Sim, redes sociais, virais	Não	tweets. Não	Sim, Twitter
A ferramenta possibilita a mineração em diferentes plataformas?	Não	Não	Sim	Não	Sim	Não
A ferramenta realiza a mineração de maneira automática?	Sim	Sim	Sim	Sim	Sim	Sim
A ferramenta realiza a mineração de forma semi-automática?	Não	Sim	Sim	Não	Sim	Não
A etapa de subjetividade é realizada de maneira automática?	Sim	Sim	Sim	Sim	Sim	Sim
A etapa de extração é realizada de maneira automática?	Sim	Sim	Sim	Sim	Sim	Sim
A etapa de Classificação é realizada de maneira automática?	Sim	Sim	Sim	Sim	Sim	Sim
A etapa de sumarização é realizada de maneira automática?	Sim	Sim	Sim	Sim	Sim	Sim
A etapa de subjetividade é realizada manualmente?	Não	Não	Não	Não	Não	Não
A etapa de extração é realizada manualmente?	Não	Não	Não	Não	Não	Não
A etapa de Classificação é realizada manualmente?	Não	Sim	Não	Não	Não, até certo ponto.	Não
A etapa de sumarização é realizada manualmente?	Não	Não	Não	Não	Não	Não
A ferramenta disponibiliza base de dados para análise?	Sim	Sim, online.	Sim	Não	Sim	Sim
Permite integração com outras mídias para busca de opiniões?	Não na versão trial	Sim, mídias sociais, virais	Não	Não	Sim, Twitter	Sim, Twitter
Possibilita a busca de opiniões em tempo real?	Não	Sim	Não	Não	Sim	Sim
É possível efetuar a análise com uma base de dados criada externamente?	Sim	Não	Sim	Sim	Sim	Não
É preciso criar uma base de dados?	Não	Não	Não	Sim	Não	Não
A base de dados da ferramenta recupera informações suficientes na etapa de	n/a	Sim	n/a	n/a	Sim	Sim

sumarização para se chegar à uma conclusão?						
Os resultados são apresentados em diferentes visualizações?	Sim, gráficos e tabelas	Não, apenas relatório	Sim, tabelas e nuvem de palavras	Sim, diferentes cores	Não, apenas relatório	Sim, gráfico e relatório
A etapa de sumarização gera um relatório que pode ser exportado?	Sim	Sim	Sim	Sim	Sim, em alguns casos	Não
O relatório pode ser exportado em pdf?	Sim	Sim	Sim	Sim	Não	n/a
O relatório pode ser exportado em MS-Word?	Sim	Não	Sim	Sim	Não	n/a
O relatório pode ser exportado em MS-Excel?	Sim	Sim	Sim	Sim	Sim	n/a
O relatório pode ser exportado em outros programas?	Sim	Sim	Sim	Sim	Sim, CSV.	n/a
É possível conectar a ferramenta às contas pessoais nas redes sociais?	Não	Sim	Não	Não	Não	Sim
É possível fazer a análise on-line com mais de um termo referente ao mesmo assunto?	Não	Sim	Não	Não	Não	Não
É possível restringir a curva de tempo de busca de um termo?	Não	Sim	Não	Não	Não	Não
A ferramenta limita o número de buscas	Sim, até o fim das linhas do MS-Excel	Sim, 1000 ocorrências na versão trial	Sim, até o fim das linhas do MS-Excel	Sim, até o fim das linhas do MS-Excel	Não	Não está explícito
É possível adicionar termo ignorado?	Sim	Sim	Sim	Não	Não	Não