

GT 8: Informação e Tecnologia

ANÁLISE DE SENTIMENTO SOBRE VEÍCULOS EM REDES SOCIAIS

Modalidade de apresentação: Comunicação Oral

Renata Maria Abrantes Baracho Porto – UFMG

Marcello Peixoto Bax – UFMG

Luiz Gustavo da Fonseca Ferreira – UFMG

Gabriel Caires Silva – UFMG

renatabaracho@ufmg.br

Resumo: Apresenta-se resultado do projeto de pesquisa que objetiva criar um processo de análise de sentimento baseado em ontologias no domínio automobilístico e desenvolver um protótipo. O processo busca analisar quais sentimentos e opiniões têm os usuários de mídias sociais na Web sobre marcas e partes de veículos de passeio. O método que orientou a elaboração do processo inclui a construção de ontologias e de um dicionário de termos que refletem a estrutura do vocabulário do domínio. O processo proposto é capaz de gerar informações que respondem a perguntas tais como: Qual carro é melhor na opinião dos consumidores, Corsa ou Palio? Qual deles é mais bonito? Qual deles possui motor 'mais potente'? Além disso, através de comparações, pode-se mostrar a opinião geral refletida em diferentes redes sociais, apontando, por exemplo, que para um dado veículo uma determinada porcentagem da opinião é considerada positiva, enquanto para outro, determinada porcentagem é considerada negativa. Os resultados podem ser usados para variados fins, como para orientar decisões de ações a serem tomadas no intuito de melhorar os produtos e/ou direcionar estratégias de marketing mais específicas. O processo pode ser generalizado e aplicado a outros domínios onde militam organizações interessadas em monitorar opiniões emitidas sobre produtos e serviços nas redes sociais.

Palavras-chave: Análise de sentimento. Redes sociais. Ontologia. Extração de informação. Recuperação de informação. Web.

1. INTRODUÇÃO

É notável o crescimento da quantidade de informação disponível na Web nos últimos anos. Páginas Web agrupam opiniões e sentimentos sobre os mais diversos objetos, tais como, resenhas de filmes, análises de produtos entre outros. Com o advento da Web 2.0, além dos conteúdos disponibilizados por empresas de notícias, os usuários passaram a compartilhar críticas e opiniões em blogs pessoais, sites de relacionamentos, microblogs, dentre outros meios.

Este artigo apresenta os primeiros resultados do projeto de pesquisa cujo objetivo principal é criar um modelo de representação do conhecimento no contexto das redes sociais (ou da web) com a elaboração de um protótipo de análise de sentimento de marca automobilística com uso de ontologias em mídias sociais. Os objetivos específicos incluem a criação de ontologias, a criação do dicionário de termos que refletem sentimentos e o desenvolvimento de um protótipo para detecção automática de sentimento.

O projeto está em desenvolvimento na linha de pesquisa de Gestão da Informação e do Conhecimento e consiste no desenvolvimento de modelo e elaboração de um protótipo de sistema de análise de sentimento em mídias sociais para o mercado automobilístico. O protótipo é aplicado à uma empresa específica desse mercado e apresenta caráter inovador de inteligência e monitoramento que combina informações de diversas fontes externas à empresa.

As informações são coletadas, classificadas e sumarizadas por meio de técnicas clássicas combinadas com tecnologias semânticas baseadas em ontologias e vocabulários. A proposta apresenta abordagem que combina o uso de ontologias em conjunto com outras técnicas da linguística, como análise de palavras e técnicas de análise morfológica, sintática e semântica. Com o uso de abordagem híbrida, espera-se gerar classificadores de sentimentos e opiniões que melhorem a acurácia dos resultados obtidos até então na área de análise de sentimento e mineração de opinião. O uso de ontologias faz com que o campo da Extração de Informação, que visa recuperar informações de textos em linguagem natural processando-as automaticamente, se relacione com a área de Representação de Conhecimento.

O resultado é o desenvolvimento do protótipo capaz de monitorar o sentimento sobre as marcas de automóveis na Web, por exemplo, o que as pessoas andam dizendo sobre o FIAT Punto nas mídias sociais? Esses resultados podem ser usados com o intuito de melhorar os produtos ou direcionar estratégias de marketing além de ser aplicável em negócios e outras organizações interessadas em monitorar opiniões emitidas sobre seus produtos e serviços.

O projeto é financiado por órgão de fomento estadual. Do grupo participam dois pesquisadores doutores e dois bolsistas de pós-graduação. A pesquisa envolve estudo em gestão da informação, extração, recuperação da informação, análise de sistemas de informação, tecnologias semânticas e filosofia da informação.

O artigo apresenta o referencial teórico com trabalhos correlatos e, em seguida, descreve o método que permitiu a criação do novo processo de análise de sentimento proposto pela pesquisa. A implementação do protótipo foi realizada para viabilizar o reconhecimento automatizado da percepção do sentimento da marca nas redes sociais. Finalmente, os resultados da aplicação do processo no domínio automobilístico são apresentados e discutidos.

2. REFERENCIAL TEÓRICO

Com a ascensão da Web 2.0 e dos portais especializados, blogs e redes sociais tornam acessíveis diariamente uma quantidade imensa de opiniões pessoais, sejam resenhas, classificações, recomendações ou mesmo outras formas de expressão online. Informações que antes eram obtidas por meio de um custoso e demorado processo de pesquisa de satisfação e opiniões, agora podem ser obtidas em larga escala na Web. O novo desafio, entretanto, está em como processar e interpretar essa grande quantidade de informações. Esse desafio é objeto de pesquisa da disciplina denominada "análise de sentimento e mineração de opiniões". Nesta pesquisa, considera-se a seguinte definição para o conceito "análise de sentimento e mineração de opiniões": a identificação, extração e o estudo de opiniões, sentimentos e emoções expressas em textos da *web*.

Segue, na próxima seção, o referencial teórico do projeto de pesquisa relatado aqui que está baseado em técnicas de análise de sentimento e mineração de opiniões com o apoio de ontologias. Os trabalhos de pesquisa revisados abaixo inspiraram os desenvolvimentos do método e do processo de análise de sentimento detalhados na Seção 3.

2.1 Ontologias

Conforme Cicortas, Iordan e Fortis (2009), há uma crescente demanda por sistemas de informação orientados para a interpretação da linguagem humana. Esses sistemas são projetados para que, com um mínimo de intervenção humana, sejam capazes de entender as intenções e

opiniões do autor. No artigo intitulado *Considerations on Construction Ontologies* os autores identificam o desafio da interpretação de informações heterogêneas por ferramentas automatizadas e analisa possibilidades de uso da ontologia para resolver essas questões. A combinação de regras ontológicas e linguagem natural são apontadas como uma solução na melhoria dos resultados da análise de sentimento. Ainda nesse contexto, foi destacada a importância das ontologias na identificação do sentido das informações, por descrição detalhada de sistemas complexos.

Rosner e Kunze (2003) discorrem a respeito das melhores práticas para construção de ontologias. Os autores apresentam suas experiências relacionadas à construção de novas ontologias, detalhando diversos métodos para o aproveitamento de restrições de linguagens, princípios de design e ideias para *frameworks*. Destacam a importância de haver um sistema de qualidade para detectar sinônimos em um processo de criação de ontologias, já que a sua ausência, em muitos casos, pode comprometer a qualidade dos resultados obtidos.

Polpinij e Ghose (2008), no artigo *An Ontology-based Sentiment Classification Methodology for Classification for Online Consumer Reviews*, apresentam como proposta uma abordagem ontológica baseada em variação léxica. Os autores propõem a utilização de três fontes para a construção de uma estrutura ontológica: um dicionário, uma lista de verbos e um conjunto de textos. A partir dessas fontes, a ontologia é construída embasada em três tipos de informações: análise morfológica, indicando um padrão na composição da palavra; análise sintática, contendo informações acerca de sua classificação (e.g. verbo) e sufixos (e.g. es); e, por fim, uma análise semântica baseada em restrições lógicas de sinonímia, antonímia e subsunção (relação “é um”). A estrutura ontológica derivada é, então, utilizada para a criação de um modelo BOW (“*bag of words*”) e alimentada em um classificador. Segundo os autores, essa técnica obteve resultados satisfatórios, chegando a atingir 96% de precisão.

Kunze e Rosner (2005) apresentam uma metodologia de extensão de ontologias usando derivação de conceitos de um domínio específico. O método utiliza uma ontologia inicial e um corpus parcialmente processado no domínio. A abordagem utilizada é baseada em estruturas sintáticas e gramaticais e, basicamente, explora características da linguagem contida no *corpus* de entrada.

2.2 Análise de Sentimentos

Bing Liu (2010) apresenta uma introdução aos principais problemas e soluções existentes no âmbito da área de pesquisa análise de sentimento, destacando a sua importância tanto para indivíduos, quanto para empresas em pesquisas de mercado, interesse e satisfação dos clientes. O texto traz definições importantes, como os conceitos de objeto e *features* (propriedades ou partes de um objeto) e opinião (sentimento positivo, negativo ou neutro em relação a um objeto ou à uma *feature*).

Wang *et. al.* (2011) propõem um método de seleção de *features* para a classificação dos sentimentos. Baseado na análise de discriminante linear (*Fisher's discriminant ratio*), o método utiliza-se do conceito de ganho de informação (*Information Gain*) e é validado a partir da comparação com outros métodos baseados nesse conceito. No artigo, os autores apresentam os resultados de dois experimentos nos quais testaram métodos de seleção de *features* diferentes. Os resultados experimentais indicaram que o método de discriminante linear possui melhor performance do que os outros analisados.

As abordagens utilizadas por Ramanathan e Ramnath (2010) exploram a utilização do contexto na análise de sentimentos por três técnicas. A primeira consiste em uma abordagem que faz uso de ontologia de domínio mapeando sentenças em objetos na ontologia. Para cada objeto, é definido um peso positivo e um negativo e a pontuação positiva e negativa de uma sentença é definida como o somatório desses pesos. Os pesos são definidos utilizando técnicas de aprendizagem de máquina e regressão. Já a segunda abordagem faz uso de uma técnica de captura de sequências de caracteres que aparecem frequentemente. Para cada par de sentenças, é extraído um conjunto de palavras que aparece em ambos e, a cada um desses conjuntos, atribui-se uma pontuação positiva ou negativa de acordo com a frequência com que aparecem nas frases. Por fim, são utilizadas duas abordagens que combinam três técnicas para a classificação de polaridade de uma sentença e os resultados são apresentados.

Wei e Gulla (2010) apresentam técnica de análise de sentimentos baseada em uma árvore ontológica de *features*. A árvore SOT (*Sentiment Ontology Tree*) é construída de modo a representar, hierárquicas, as *features* de um objeto. Cada nó da árvore contém como filhos, além das *features* que o compõem, dois nós folha, representando os sentimentos negativo e positivo do *feature* representado pelo nó. A abordagem de classificação utilizada é baseada em algoritmo de

classificação hierárquica. O algoritmo recebe como parâmetros de entrada uma SOT e textos já classificados e tem como objetivo validar a construção hierárquica sentimental dos textos. Os resultados obtidos foram capazes de demonstrar que o conhecimento sobre as relações hierárquicas melhoram o desempenho e precisão da análise de sentimentos. Além disso, é possível utilizar um modelo genérico com uma SOT composto de SOTs de objetos individuais e um nó raiz. Essa adaptação permite que o algoritmo seja utilizado em textos genéricos, ou seja, que não contenham um objeto predefinido.

Mosha e Tianfang (2010) apresentam uma metodologia para identificar a relação opinião “elemento-de-referência” em passagens de texto utilizando análise de *"parsing"* de dependências, assim como análise semântica superficial. Essa análise é feita com base em uma ontologia de palavras (por dicionário) e um banco de dados de posicionamento de termos (*collocation*) buscando considerar toda a semântica por trás do tópico e do sentimento. O algoritmo utilizado no *parsing* segue basicamente quatro passos: o primeiro identifica o objeto da frase (sujeito nominal, objeto direto ou um sujeito oculto dado numa modificação adverbial, comum no chinês); o segundo detecta relações lógicas com os objetos (detectando coisas como *features* ocultas); os dois últimos passos são basicamente heurísticas para identificar relações indiretas. A cada um dos passos, fragmentos identificados contendo sentimentos vão sendo relacionados com objetos da ontologia. Os resultados mostram que a técnica pode ser usada em *corpus* genéricos (como o COAE2008, *corpus* público chinês para avaliação de abordagens de análises de sentimento).

Neviarouskaya, Prendinger e Ishizuka (2011), no artigo *A Lexicon for Sentiment Analysis*, descrevem um método para automatização da geração e marcação de valores de nível de sentimento para fragmentos de texto subjetivos chamado *SentiFul*. A ideia é possibilitar expandir uma base qualquer por técnicas como sinonímia direta, aproveitamento de relações de antonímia, hiponímia derivação, composição, entre outros. Na proposta, é feito um reconhecimento nos termos textuais de quatro tipos de afixos (usados na derivação de novas palavras), dependendo do papel que desempenham como: propagação, reversão, intensificação, e enfraquecimento. Além de derivação, é feito o processo de encontrar novas palavras usando composição, um processo que gera um grande número de termos úteis, especialmente no caso dos substantivos e adjetivos. O algoritmo foi elaborado para fazer extração automática de palavras relacionadas com sentimento, utilizando termos do *WordNet* (mas usando palavras da *SentiFul*).

3. FERRAMENTAS E MÉTODOS

Essa seção apresenta as ferramentas e o método que foram utilizados no processo de análise de sentimento e, em seguida, a descrição detalhada do processo elaborado e proposto nesta pesquisa.

3.1 Ferramentas para o processo de análise de sentimento

A seguir, serão descritos o software PALAVRAS, que faz a análise semântica de texto e foi utilizado no desenvolvimento do processo, e criado o dicionário de sentimentos.

3.1.1 O software Palavras

O processo desenvolvido na pesquisa consistiu, primeiramente, em analisar semanticamente fragmentos de textos (artigos e comentários) de redes sociais, com objetivo de extrair informações de sentimentos. Para se fazer a análise semântica do texto, foi utilizado o software PALAVRAS (desenvolvido por *Eckhard Bick* e baseado no corpus "Floresta Sintática" da Linguatca). Trata-se de um analisador sintático automático para o português que realiza análise gramatical e sintática da língua portuguesa e é capaz de fornecer informações morfológicas de uma frase.

O processo de análise de sentimento elaborado pela pesquisa inicia-se com a utilização do software PALAVRAS como analisador sintático e léxico. Esse software é utilizado na base do algoritmo, como "normalizador" da entrada e analisador sintático.

O processo da análise de sentimento começa com a extração de elementos do texto relacionados à opinião. Em seguida, é realizada a classificação da opinião quanto ao seu caráter, considerado dentro do escopo de positivo, negativo ou neutro. Na sequência, é realizada a comparação de sentenças quanto as suas opiniões. Comumente usa-se o termo "objeto" para referir-se ao alvo da opinião, podendo este conter diversas características ou subpartes de maneira que essas possam também ser alvo de opiniões.

3.1.2 O dicionário de sentimentos

A construção do dicionário de sentimentos foi baseada na classificação de sentimentos do dicionário *Sentistrength* por THELWAL *et. al.* (2011) um dicionário com dezenas de milhares de termos que denotam sentimento. O intuito do dicionário é quantificar sentimentos. Um exemplo de entrada nesse dicionário seria “*bad*: -2”, significa que o termo inglês *bad* carrega um sentimento negativo de valor numérico (-2). Segundo os autores, o dicionário foi construído a partir de pesquisas nas áreas de psicologia, filosofia e linguística. O maior desafio para a utilização do dicionário *SentiStrength* no projeto foi realizar o processo de tradução mantendo o real significado dos termos na língua portuguesa.

O processo de tradução do dicionário *Sentistrength* que contém aproximadamente uma lista de mais de 22 mil termos, por um processo semiautomático de tradução, foi dividido em três etapas descritas a seguir.

A primeira etapa engloba a tradução inicial. Foram utilizadas três ferramentas: o *Bing Translator da Microsoft*, o *Google Translator* e o *Yahoo BabelFish*. A partir da tradução, feita por cada uma das ferramentas, foi criado um Índice de concordância. Com ele, foi possível filtrar os termos que apresentaram maior discordância entre os dicionários e, portanto, dar-lhes maior atenção.

A segunda etapa consiste na validação da tradução dos termos. Apesar da inquestionável utilidade dos tradutores automatizados, muitos detalhes característicos de cada língua passam despercebidas, sendo necessária a intervenção manual de uma pessoa especializada em traduções. Como seria muito dispendiosa a alocação de um especialista para traduzir todos os termos em questão, decidiu-se automatizar o processo. O processo consiste na execução de um programa feito para acessar o COMPARA¹, que é um corpus paralelo Inglês/Português disponibilizado pela Linguatca (centro de recursos distribuído para a língua portuguesa). O funcionamento do COMPARA é o seguinte: dado um termo em inglês (ou português), ele nos mostra como foi feita a tradução daquele termo em vários contextos distintos, incluindo obras de Machado de Assis, Eça de Queirós e Aluísio Azevedo. O programa desenvolvido nessa etapa funciona como um

¹ COMPARA é um corpus paralelo bidireccional de português e inglês. Disponível em <http://www.linguatca.pt/COMPARA/>

*crawler*², consultando o COMPARA para cada um dos termos e registrando os casos onde as traduções são correspondentes e os casos em que não há correspondência. Dessa forma, validamos as sugestões do trabalho dos tradutores automatizados a partir de traduções feitas por profissionais especializados.

A terceira etapa foi a finalização. Nessa etapa, selecionamos os termos mais relevantes, ou seja, os termos onde traduções imprecisas resultariam num impacto negativo maior nos resultados da pesquisa. Selecionados esses termos, foi feita uma inspeção em cada uma das traduções, procurando por imperfeições no processo de tradução feito até aqui, etapa em que foram analisados cerca de 600 termos.

3.1.3 As ontologias específicas de domínio

Após a definição da análise dos textos e do processo de tradução do dicionários de termos referentes a sentimentos que serão considerados na pesquisa, passa para a etapa de definição das ontologias. O conceito de ontologias foi utilizado para guiar o processo de identificação de objetos (no caso os modelos de carro da Fiat e não Fiat), características (propriedades desses objetos como potência, beleza, entre outros), contextos (sequência de parágrafos com opiniões sobre o mesmo objeto), também são vitais no processo final de classificação. Para o desenvolvimento dessa pesquisa, empregamos o conceito de Ontologia utilizado no campo da ciência da informação como um conhecimento formal, um conjunto de conceitos e seus relacionamentos em um domínio. Ontologias podem ser utilizadas para modelar conceitos humanos e abstratos. A formalização permite que sistemas possam tirar proveito de modelos humanos.

Foram criadas ontologias baseadas nos modelos da FIAT. A primeira ontologia "Modelos da FIAT" utilizada na pesquisa agrega as informações necessárias sobre os modelos dos automóveis da Fiat. Essa ontologia nos permite não apenas detectar os objetos de interesse no texto (como "Fiat Palio" por exemplo) mas também a classe que o dado objeto pertence (um automóvel "Hatch" do subtipo "Hatch Compacto" conforme Figura 1.

² *crawler* é um programa de computador que navega pela web de uma forma metódica e automatizada; os motores de busca, usam *crawler* para manter uma base de dados atualizada

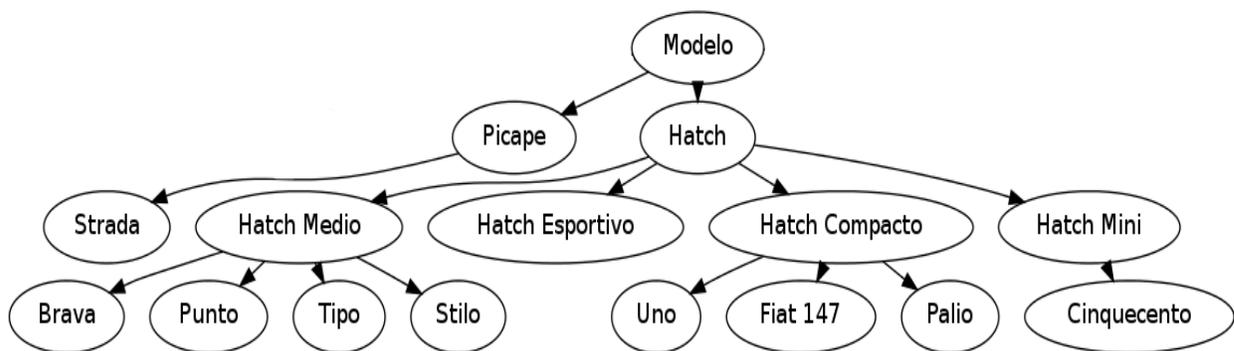


Figura 1 – Exemplo de parte da ontologia: Modelos da FIAT
 Fonte: Elaborado pelo autor.

A segunda ontologia "*Features* de automóvel" organiza as características como relevantes em um automóvel, ou seja, agrega componentes e características que são alvos de sentimentos (Figura 2).

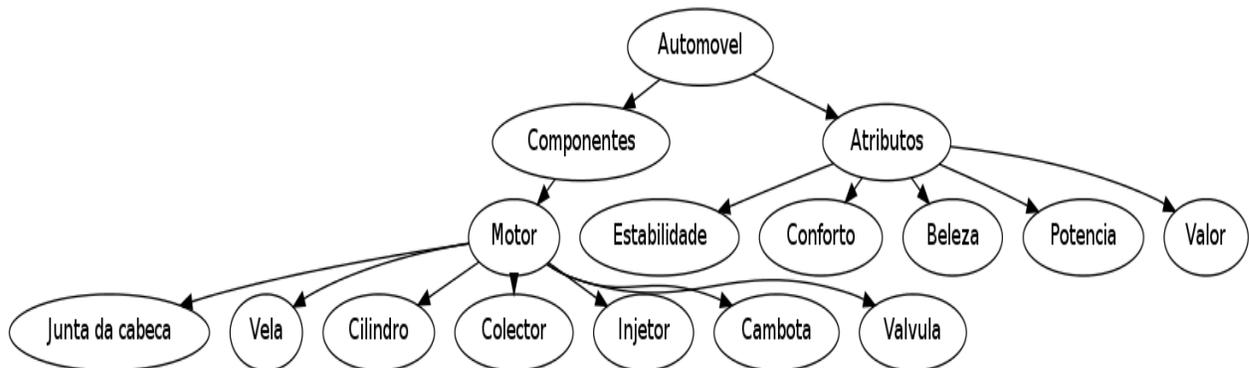


Figura 2 – Exemplo de parte da ontologia: *Features* de automóveis
 Fonte: Elaborado pelo autor.

A terceira ontologia "Modelos Não-FIAT" é formada por outros objetos (automóveis) que possuam as mesmas *features* (características e componentes) dos objetos de interesse. As *features* ligadas a esses objetos não-FIAT devem ser detectadas e corretamente descartadas da análise, pois causariam distorções do resultado (Figura 3).

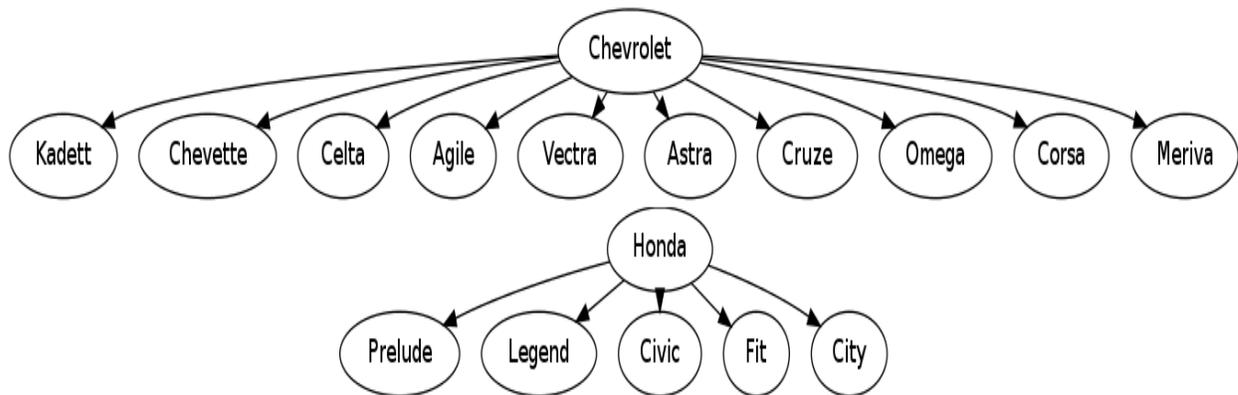


Figura 3 – Exemplo de parte da ontologia: Modelos Não FIAT
Fonte: Elaborado pelo autor.

3.2 Visão completa do processo de análise

Após a definição das ontologias, o processo de análise de sentimento proposto foi implementado em um protótipo. Genericamente, o processo consiste na captação, análise e armazenamento das opiniões. Mais especificamente, o processo foi dividido em oito etapas que contemplam desde a captação das opiniões em redes sociais até a agregação das opiniões processadas, conforme ilustra a Figura 4.

As primeiras etapas representam como o texto contendo as opiniões é capturado e normalizado. Em seguida, os objetos de interesse (definidos pela ontologia) são encontrados no texto, assim como suas *features* (características). Posteriormente, é feita a detecção e o cálculo (com base no dicionário de sentimentos) dos sentimentos relacionados a cada um dos objetos de

interesse e suas *features*. Por fim, os resultados são analisados e armazenados em forma de relatórios.

A seguir são detalhadas as etapas ilustradas na Figura 4.

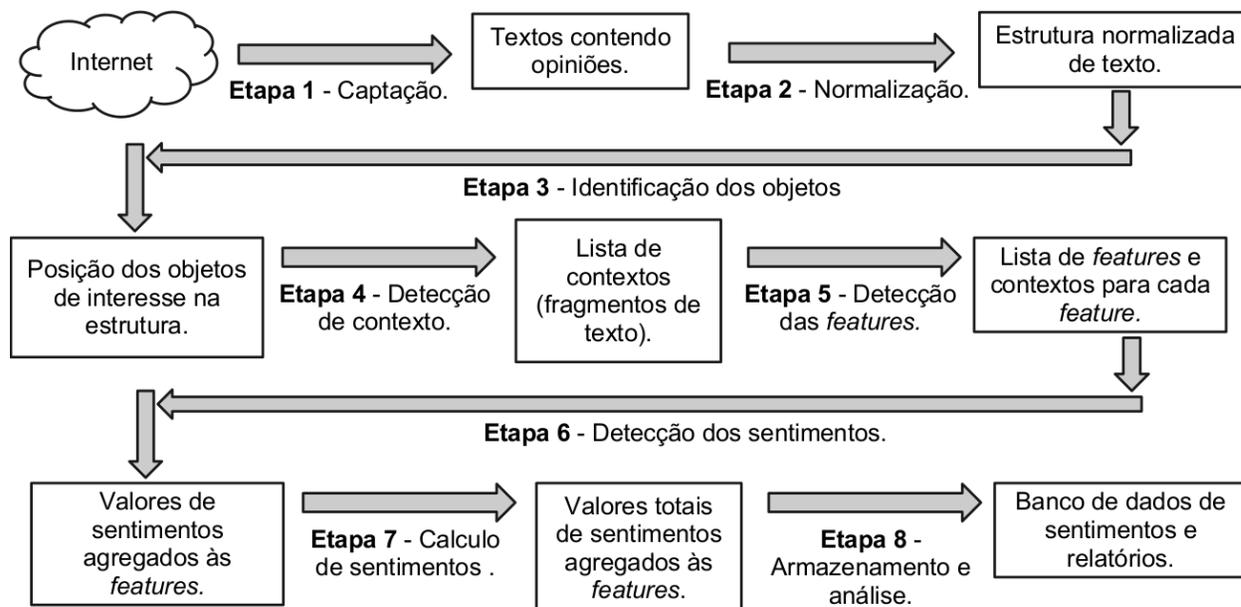


Figura 4 – Visão geral do processo de análise de sentimentos proposto pela pesquisa.

Etapa 1 - Captação

Inicialmente, vários blogs com a temática "automóveis" e que permitem a discussão aberta entre os leitores foram acessados por um *crawler* para construir a base de textos (*corpus* analisado). Um programa de captação foi desenvolvido para acompanhar novos artigos e notícias dos sites selecionados por RSS (*RDF Site Summary*). O processo de captação iniciado pela seleção de *RSS Feeds* (agregador das notícias em RSS) de sites de interesse, como portais e blogs com o tema automóveis que então o conteúdo interno, é extraído e armazenado como dados XML e HTML brutos (formatos em que RSS são armazenados na WEB). Em uma segunda etapa, essas informações brutas são processadas separando o código HTML e XML dos textos de interesse da notícia e dos eventuais comentários. O resultado desse passo é uma coleção de textos agrupados como notícia/artigo (título e texto) e os textos de quaisquer comentários relativos. Esse *corpus* compõe a entrada para o algoritmo de normalização.

Etapa 2 – Normalização

Após a captação dos textos, foi necessário normalizá-lo a fim de prepará-lo para análise pelo algoritmo. Para isso é necessário obter a estrutura léxica e morfológica dos textos, além de trazer todos termos para sua forma mais básica (infinitivo). Esse passo é realizado pelo software PALAVRAS. Exemplo de normalização utilizando o palavras em um texto capturado (Figura 5): "Não sei se é o melhor, mas estou muito satisfeito com o meu Pálio. É econômico e nunca me deu problemas e já o possuo desde 2003." Texto retirado de <http://br.answers.yahoo.com/question/index?qid=20061112175518AAfs7a3>.

```
não [nã] ADV @ADVL>
sei [saber] <vt> <fmc> V PR 1S IND VFIN
@FMV
se [se] KS @SUB @#FS-<ACC
é [ser] <vK> V PR 3S IND VFIN @FMV
o [o] <artd> DET M S @>N
melhor [bom] <KOMP> <SUP> <n> ADJ M
S @<SC
,
mas [mas] KC @CO
estou [estar] <vK> <fmc> V PR 1S IND
VFIN @FMV
muito [muito] <quant> ADV @>A
satisfeito [satisfazer] <vt> V PCP M S
@<SC
com [com] PRP @<ADVL
o [o] <artd> DET M S @>N
meu [meu] <poss 1S> DET M S @>N
pálio [pálio] N M S @P<
```

Figura 5 – Codificação da Normalização

Etapa 3 - Identificação de objetos de interesse.

Com os textos normalizados, nesta etapa as ontologias dos automóveis da marca FIAT (além da própria palavra FIAT) e a ontologia das marcas "não-FIAT" são utilizados para realizar-se a detecção dos objetos de interesse. Essa etapa tem como resultado as posições das palavras no textos identificadas como objetos de interesse, sendo os objetos os modelos de carros identificados tanto FIAT (modelos de carro identificados da Fiat) quanto não-FIAT (modelos de outras empresas), chamamos essas posições simplesmente de marcadores de objetos que representam a posição das palavras no texto, identificadas como objetos de interesse (os modelos

dos carros). A contabilização do posicionamento é feita considerando a quantidade de palavras necessárias para se chegar à palavra do objeto dado início da frase ou parte do texto. Ao ler o texto, a primeira palavra tem a posição zero; a segunda, a posição um, e assim sucessivamente.

Por exemplo: "*Eu prefiro o Corsa, é mais macio e confortável. O Pálio também é bom e tem mais espaço interno. Já o Gol, acho muito duro.*" retirada de <http://br.answers.yahoo.com/question/index?qid=20070227083007AAf9IKv>. Esse texto possui um objeto de interesse presente na ontologia FIAT (o objeto Pálio) na posição 10, e outros dois objetos da ontologia não-FIAT (os objetos Corsa posição 3 e Gol posição 21).

Etapa 4 - Detecção de contexto

A próxima etapa consiste na extração dos contextos dos objetos detectados na etapa anterior. Um contexto, no processo proposto, é representado por um fragmento de texto. Assim, o texto da frase exemplo da Etapa 3 seria dividido em contextos conforme mostra a figura abaixo (Figura 6).

Eu prefiro o Corsa, é mais macio e
"confortável". O Pálio também é bom, e tem
mais espaço interno. Já o Gol, acho muito
duro.

Figura 6 – Detecção de contexto

Etapa 5 - Detecção das características (*features*)

Separados os contextos, na próxima etapa acontece a detecção das características ou subpartes (*features*) dos automóveis em cada um dos contextos gerados na etapa anterior. Nesse caso utiliza-se a ontologia de *features* de carros (Figura 2) para guiar o processo (Figura 7).

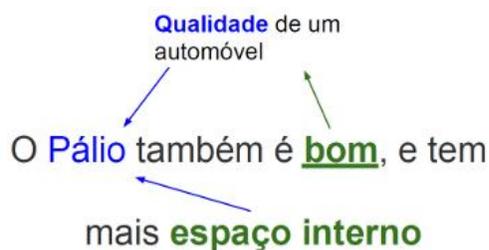


Figura 7 – Detecção de features

Etapa 6 - Detecção de sentimentos

Após a detecção das *features*, é feito o processo de detecção de sentimentos. Nesse estágio se utiliza o *SentiStrenght* para detectar e classificar o nível de sentimento. Sentimentos são ligados às *features* mais próximas (Figura 8).

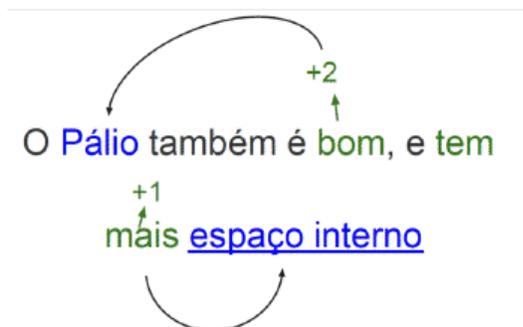


Figura 8 – Detecção de sentimentos

Etapa 7 - Agregando sentimentos

Em seguida é feita uma agregação dos sentimentos detectados à estrutura da ontologia de *features* de carros (Figura 2). Pode-se assim inferir, por exemplo, que se alguém fala bem da potência de um carro, ele está a se referir indiretamente ao motor do mesmo. Em uma consulta do sentimento dado um ponto qualquer da árvore de ontologia, os valores de sentimento agregados aquele ponto à descendentes são também contabilizados para o pai, assim como qualquer ponto ascendente ao atual contabiliza os sentimentos do ponto atual. Isso permite contabilizar de forma que um sentimento positivo agregado aos pneus de um modelo de carro, automaticamente, seja contabilizado ao próprio carro nessa agregação, chegando até a marca do carro que recebe todos os sentimentos na raiz.

Etapa 8: Armazenamento e Análise

Nessa etapa, todas as informações de sentimento são armazenadas, mantendo as referências de modelo do carro (ontologia FIAT), *feature* (ontologia automóveis) e contextos para validações futuras. Isso permite diversos tipos de referências cruzadas e comparações.

4. RESULTADOS

Um dos resultados obtido nesta pesquisa foi a construção das ontologias de objetos de interesse da marca de automóvel pesquisada (Fiat), dos objetos não-Fiat e dos componentes de carros. Essas ontologias representam os conceitos e relações dentro do domínio do mercado de carros. A partir desses modelos de dados, é possível realizar inferências sobre os objetos do domínio. Foram utilizadas tanto para identificação dos objetos, quanto para identificação das características desses objetos, possibilitando ao sistema inferir sobre uma determinada opinião.

A pesquisa demonstra que a metodologia desenvolvida e a estrutura para executá-la se provou funcional. Um resultado é a prova de conceito que é possível minerar as redes sócias para gerar estatísticas sobre a opinião e o sentimento das pessoas na Web.

O protótipo desenvolvido permite analisar as opiniões relacionadas a um determinado objeto da ontologia e suas características. A partir dessas informações, é possível gerar inúmeros tipos de relatórios. Um exemplo são os relatórios de estudos comparativos. A partir deles, é possível responder a perguntas como: Qual carro é melhor na opinião dos clientes, Corsa ou Palio? Qual deles é mais bonito? Qual deles, ainda segundo a opinião dos clientes, possui um motor “mais forte”?

A Figura 9 apresenta um modelo de saída de relatório que se pode obter a partir do protótipo. Como resultado nesse caso específico, extraindo as opiniões relativas aos carros Corsa, Gol e Palio nas redes sociais foi possível inferir automaticamente que existem mais opiniões positivas do que negativas sobre os carros. Em relação ao componente motor, para o carro Palio, 63% da análise de opinião é considerada positiva, enquanto 37% é considerada negativa. Para o carro Gol, 47% da análise de opinião é considerada positiva, enquanto 53% é considerada

negativa. Para o carro Corsa, 38% da análise de opinião é considerada positiva, enquanto 62% é considerada negativa.

A Figura 9 representa uma forma de apresentar os resultados.

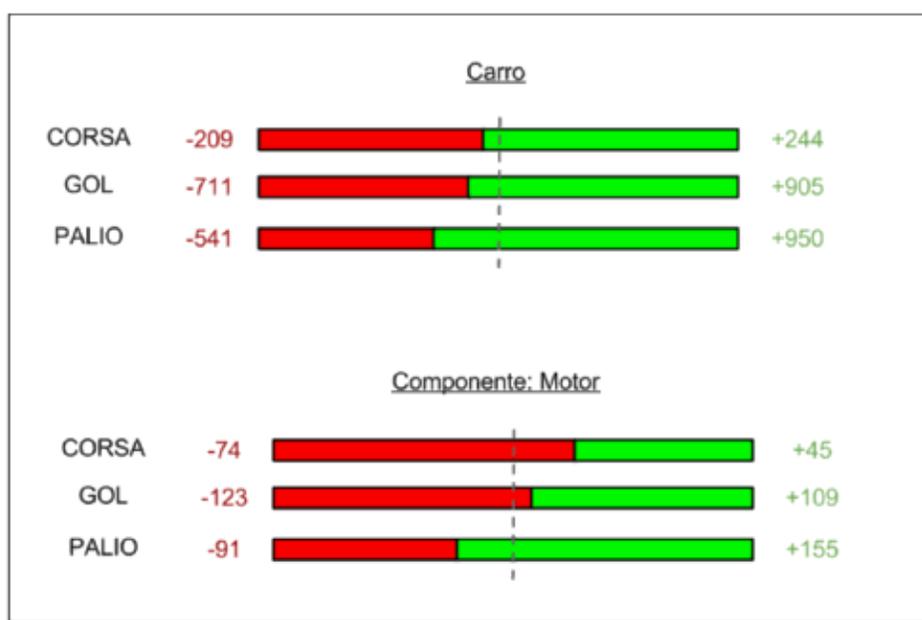


Figura 9 – Resultados parciais.

5. CONSIDERAÇÕES FINAIS

O artigo apresenta os resultados do projeto de pesquisa na área de análise de sentimentos e mineração de opiniões na Web 2.0. Foi desenvolvido um processo de análise de sentimento, e criado um processo próprio que, por hora, foi aplicado ao domínio automobilístico e poderá ser generalizado e aplicado a outros domínios.

Preliminarmente, apenas a título de exemplo do tipo de informação que se pode obter do sistema no domínio de veículos, pôde-se concluir com o uso do protótipo, como mostra o resultado da Figura 9, que existem mais pessoas falando bem do que mal do automóvel FIAT

Palio, entretanto, existem mais pessoas falando mal do Corsa e do Gol do que falando bem. Além disso, pode-se comparar a opinião dos usuários em relação aos três carros. Trata-se de um resultado automático construído pelas opiniões que circulam nas redes sociais sobre tais veículos.

Através da aplicação da ontologia no modelo podemos inferir e comparar parte do carro (motor, carroceria) em diversas marcas ou comparar por subtítulo (Fiat e não Fiat) quem está melhor em uma determinada categoria.

Em junho de 2012, foi realizada uma apresentação parcial do projeto. A partir das impressões dos representantes da empresa que foram compartilhadas com o grupo, foi possível perceber o entusiasmo e o impacto que a ferramenta pode causar no departamento de gestão e marketing da empresa. Com efeito, entender como um produto é visto por clientes é o primeiro passo para melhorar sua imagem.

O grupo foi questionado a respeito das possibilidades do processo descrito na Seção 4: coletar informações a respeito do autor da opinião e apresentar relatórios em uma linha do tempo, de forma que o responsável pela análise pudesse identificar tendências. Tais questões foram consideradas para desenvolvimentos futuros que enriquecerão o processo aqui descrito, pois a possibilidade de identificar perfis e analisar tendências de opiniões dos clientes abrirá inúmeras possibilidades de aplicação desta pesquisa. Pretende-se generalizar o processo proposto para que possa ser aplicado a outros domínios e organizações interessadas em monitorar opiniões emitidas sobre produtos e serviços.

REFERÊNCIAS

CICORTAS, A.; IORDAN, V. S.; FORTIS, A. E; Considerations on Construction Ontologies. **Annals Computer Science Series**, Timișoara, Romania, v. 7, n. 1, p. 79-88, mai 2009.

KUNZE, M; ROSNER, D.; **Context-related Derivation of Word Senses**. In: INTERNATIONAL LANGUAGE RESOURCES AND EVALUATION CONFERENCE, 4, 2004, Lisboa, Portugal. **Proceeding of the Ontologies and Lexical Knowledge Bases Workshop....** Paris: European Language Resources Association, 2004

LIU, B. Sentiment Analysis and Subjectivity. In: INDURKHYA, Nitin; DAMERAU, Fred J. (Orgs.) **Handbook of Natural Language Processing**, 2. nd. Boca Raton: CRC Press, 2010. p. 627-666

MOSHA, C.; TIANFANG, Y. Combining Dependency Parsing with Shallow Semantic Analysis for Chinese Opinion-Element Relation Identification. In: INTERNATIONAL UNIVERSAL

COMMUNICATION SYMPOSIUM, 4., 2010, Beijing, China. **Proceedings of IUCS...** [S.l. : s.n], 2010

NEVIAROUSKAYA, A.; PRENDINGER, H.; ISHIZUKA, M. SentiFul: a Lexicon for Sentiment Analysis. **IEEE Transactions on Affective Computing**, Washington, DC, v. 2, n.1, p. 22-36, jan. 2011.

POLPINIJ, J.; GHOSE, A. K. An Ontology-based Sentiment Classification Methodology for Classification for Online Customer Reviews. In: THE 2008 IEEE/WIC/ACM INTERNATIONAL CONFERENCE ON WEB INTELLIGENCE AND INTELLIGENT AGENT TECHNOLOGY, 2008, Sydney, Australia.[**Trabalhos apresentados...**] [S. l.: s.n], 2008, p. 518-524.

RAMANATHAN, J.; RAMNATH, R. **Context-assisted Sentiment Analysis**. Columbus, Ohio: Ohio State University, 2010. Disponível em: <<http://ceti.cse.ohio-state.edu/architectures-services/technical-reports/OSU-CISRC-6-09-TR28%20-%20Context-Assisted%20Sentiment%20Analysis.pdf/view>>. Acesso em: 10 jun 2012.

ROSNER, D.; KUNZE, M.; Exploiting Sublanguage and Domain Characteristics in a Bootstrapping Approach to Lexicon and Ontology. In: INTERNATIONAL LANGUAGE RESOURCES AND EVALUATION CONFERENCE, 3., 2002, Las Palmas, Spain. **Proceeding of the Ontologies and Lexical Knowledge Bases Workshoph...** Ontolex2002. Paris: European Language Resources Association, 2003, p. 68-73

THELWALL, M.; BUCKLEY, K.; PALTOGLOU, G. Sentiment Strength Detection for the SocialWeb. **Journal of the American Society for Information Science and Technology**, v. 63, n. 1, p. 163-173, ago 2011.

WANG, S. *et. al.* A Feature Selection Method Based on Improved Fisher's Discriminant Ratio for Text Sentiment Classification. **Expert Systems with Applications**, Tarrytown, NY, v. 38, n. 7, p 8696–8702, jul. 2011

WEI, W.; GULLA, J. A. Sentiment Learning on Product Reviews via Sentiment Ontology Tree. ANNUAL MEETING OF THE ASSOCIATION FOR COMPUTATIONAL LINGUISTICS, 48., 2010, Uppsala, Sweden. **Proceedings...** Stroudsburg: Association for Computational Linguistics, 2010, p. 404-413.

Abstract: This paper presents partial results of a research project that aims to create a process of sentiment analysis based on ontologies in automobile domain and develops a prototype. The process aims to make social media analysis identifying feelings and opinions about brands and vehicle parts. The method that guided the development process involves the construction of ontologies and a dictionary of terms that reflect the structure of the vocabulary domain. The proposed process is capable of generating information that answer questions like "in the opinion of the customer, which car is better, Corsa or Palio? Which one is more beautiful? What engine is stronger? To answer these questions, by comparison, one can show a general view reflected on different social networks, indicating for example that for a given vehicle, a certain percentage of

view is considered positive, while for others, a certain percentage is considered negative. The results can be used for various purposes such as to guide decisions in order to improve the products or direct specific marketing strategies. The process can be generalized and applied to other areas in which organizations are interested in monitoring views expressed about their products and services.

Palavras-chave: Análise de sentimento. Redes sociais. Ontologia. Extração de informação. Recuperação de informação. Web.

Keywords: Sentiment analysis. Social Networks. Ontology. Extraction of information. Information Retrieval. Web.