

IDENTIFICANDO EMOCÕES EM REDES SOCIAIS: Um estudo de caso no facebook

IDENTIFYING EMOTIONS IN SOCIAL NETWORKS: a case study on facebook

SILVA, William¹; RIBEIRO, Richardson²; TEIXEIRA, Marcelo³; ENEMBRECK Fabricio⁴

¹ Acadêmico do Curso de Engenharia de Computação da UTFPR, Câmpus Pato Branco
williamsousadasilva@gmail.com

² Docente pesquisador do Departamento de Informática da UTFPR, Câmpus Pato Branco
prof.richard@gmail.com

³ Professor Adjunto da UTFPR, Câmpus Pato Branco
marcelosolidade@gmail.com

⁴ Professor Titular da Pontifícia Universidade Católica do Paraná e
Pesquisador do Programa de Pós-Graduação em Informática da PUC-PR
fenembreck@gmail.com

Resumo

Neste trabalho é apresentado um modelo de reputação para identificar o nível emocional de indivíduos a partir de conteúdos postados em redes sociais. A proposta do modelo é caracterizar um indivíduo emocionalmente, como forma de estimar o quão confiável um indivíduo pode estar em problemas de tomada de decisão, seja coletivo ou individual. Essa abordagem tem sido pouco explorada na literatura e, quando o é, foca apenas na avaliação de produtos e serviços a partir de opiniões de usuários. Os resultados desse trabalho são ilustrados por meio de um exemplo que identifica e compara informações compartilhadas por usuários em redes sociais, com um conjunto de palavras (base textual) afetivas, definidas pelo Anew-br como forma de detectar sentimentos e caracterizar um indivíduo emocionalmente.

Palavras-chave: modelo de reputação; mineração de texto; redes sociais.

Abstract

In this paper a reputation model has been presented to identify the emotional level of an individual using content posted on their social networks. The proposed model aims to emotionally characterize an individual in order to establish how confident he is for a decision-making process, both collectively or individually. This approach has been slightly exploited in the literature and, in most cases, it is used only as a way to evaluate products and services from user reviews. The results of this work are illustrated by an example that identifies and compares information shared by users in social networks against a set of words (text-based) defined by the Anew-br, as a way to detect feelings and so emotionally characterize an individual.

Key-words: reputation model; text mining; social networks

INTRODUÇÃO

A ampla aceitação das redes sociais e sua extensiva demanda de usuários tem proporcionado à comunidade científica realizar pesquisas nas mais diversas áreas do conhecimento, como psicologia comportamental (Wasserman e Galaskiewicz, 1994), marketing (Weber, 2009), sociologia (Wasserman e Galaskiewicz, 1994), ciência da computação (Alt et al., 2006), etc.

Uma das formas de se implementar esse tipo de pesquisa é por meio dos modelos de reputação e confiança, criados com o intuito de identificar a forma como os indivíduos se comportam na rede social, ou seja, a forma como se integram ao ambiente e aos outros indivíduos.

No contexto da ciência da computação, um modelo de confiança e reputação permite identificar, por exemplo, se uma informação recebida de outro indivíduo é ou não confiável (Sabater and Sierra, 2001). Com base nesse tipo de estimativa, pode-se optar por aceitar ou não tal informação, além de decidir sobre o grau de atenção dedicado a ela.

Na literatura, um modelo de reputação é caracterizado por duas dimensões (Mui, 2002):

a) em nível local, no qual o grau de confiança atribuído por um indivíduo a outro indivíduo é medido através das interações e da troca de informações entre eles; e

b) em nível global, no qual a reputação é atribuída não apenas por um único indivíduo, mas pela comunidade, ou seja, a reputação é derivada a partir de conjuntos de valores de confiança, transmitidos

de indivíduo para indivíduo através de votações e testemunhos.

Um modelo de reputação também pode refletir o estado mental de um indivíduo, que está diretamente ligado à forma como suas decisões são tomadas e de algum modo refletem o seu estado emocional.

Neste trabalho é apresentado um modelo de reputação que explora a dimensão emocional de um indivíduo identificada a partir dos conteúdos postados em seus sites de relacionamento, em particular, a rede social Facebook (Facebook, 2014). Inicialmente, um conjunto representativo de informações é obtido da rede social e, em seguida, ele é comparado com outro conjunto, designado pela Anew-br como meio de detectar sentimentos e caracterizar níveis emocionais.

Este trabalho está estruturado da seguinte maneira: Na seção 2 são apresentados conceitos relacionados sobre reputação e confiança, bem como alguns trabalhos dedicados ao uso de abordagens para a identificação de níveis emocionais. Na Seção 3 é mostrado a metodologia do trabalho em questão, apresentando as etapas e o modelo de reputação proposto. Já na Seção 4 resultados experimentais são apresentados e discutidos. Conclusões são descritas na Seção 5.

FUNDAMENTAÇÃO TEÓRICA

Sabater and Sierra (2001) definem reputação como sendo a opinião de um indivíduo sobre algo, podendo ser representada de três formas: individual,

social ou ontológica. A reputação individual se refere a maneira em que um indivíduo é julgado pelos outros. A reputação social refere-se à forma com que vários indivíduos são julgados, com base no grupo social ao qual pertencem. Já a reputação ontológica depende do ambiente social em que um indivíduo se encontra, ou seja, sua reputação pode ter diferentes classificações em comunidades distintas.

A partir dessas definições podemos formalizar o conceito de modelo de reputação. Um modelo de reputação consiste em um sistema que procura gerenciar as opiniões sobre o comportamento dos indivíduos numa determinada rede e qualificar os dados envolvendo esses indivíduos. Através da combinação dessas opiniões e dos dados qualitativos que as relacionam, é possível formar a reputação, e posteriormente utilizá-la para determinar a confiança do indivíduo.

De acordo com Pinyol (2013) a confiança implica na decisão de confiar em alguém. De acordo com Bromley (1993) reputação é uma ferramenta social utilizada para reduzir a incerteza de interagir com indivíduos cujas ações e atributos são desconhecidos. Para Josang (2007) o conceito de reputação e confiança estão estreitamente ligados, visto que o primeiro é utilizado para derivar o segundo.

Os conceitos de confiança e reputação podem ser estendidos para a caracterização emocional de um indivíduo, estimando-se, assim, o quão confiável ele está, com base no seu nível

emocional. Dosciatti, et al. (2012) propuseram um método de identificação de emoções em bases textuais de língua portuguesa baseado em Latent-Semantic-Analysis. A técnica proposta foi aplicada em textos reduzidos de manchetes e de notícias extraídas de sites da internet.

Com o objetivo de identificar as emoções presentes no texto, em alguns experimentos, o método obteve uma taxa média de identificação de emoções na ordem de 70%. Já Santos (2010) desenvolveu um protótipo capaz de realizar mineração de opinião em textos de redes sociais, tendo como estudo de caso o Twitter.

O autor utilizou a linguagem Java para a implementação do protótipo e usou o método de aprendizagem de máquina SVM (Support Vector Machine) para realizar uma classificação binária de sentimentos positivos e negativos.

Para representar os documentos de texto de maneira estruturada, o autor utilizou o modelo conhecido como vetor de características, uma abordagem estatística para análise de textos. Os experimentos mostram que o classificador alcançou uma taxa média de 80% de acerto.

Esses são apenas alguns dentre vários trabalhos que se dedicam ao estudo de abordagens para mineração de texto e análise de sentimentos usando algum modelo de reputação.

METODOLOGIA

A primeira etapa para a construção do modelo proposto nesse artigo consiste em extrair

os dados a partir de textos publicados por usuários no Facebook. O Facebook é uma rede social que permite aos usuários publicarem diversos tipos de conteúdo, dentre eles: fotos, vídeos, mensagens de status, etc. Este trabalho é centrado na exploração dos conteúdos das mensagens de status, pois é nesse tipo de publicação que geralmente as pessoas expressam suas opiniões e sentimentos sobre um determinado assunto. O Facebook distribui uma interface para a camada de programação, denominada GraphAPI, que permite a integração do conteúdo da rede social com outros softwares sem, para isso, envolver-se com as suas questões de implementação. Assim, utilizando-se do software Pentaho Data Integration (PDI) (Pentaho, 2013), foi conduzido um processo de extração de dados do Facebook, como mostra a Figura 1. O PDI é um software de código aberto que pode ser utilizado para a extração, transformação e carga de dados.

Figura 1- Processo de extração de dados
Fonte: Pentaho, 2013



O processo de extração é consistido de três etapas. Na primeira (Generate Rows) é inserida uma chave gerada por um aplicativo da rede social. Na segunda etapa (HTTP client) especifica-se o local onde foi feita a extração das informações. É nessa etapa que ocorre a integração entre a API do Facebook e o Pentaho. Já na terceira etapa (Text File Output) o Pentaho é configurado para retornar

os dados em arquivos. Após isso, etapas de limpeza, seleção e transformação dos dados são conduzidas, para só então ser aplicado o processo de mineração. Essas etapas são conhecidas como Knowledge-Discovery in Databases (KDD) (Fayyad, 1996).

Após a extração dos dados, a próxima atividade do modelo consiste em comparar os conteúdos das mensagens postadas pelos usuários no Facebook com uma lista de palavras afetivas, definidas pelo Anew-br. O Anew-br (Brazilian norms for the affective norms for english words) é uma base textual com mais de 1.000 palavras em que cada uma dessas palavras possui dois valores, sendo um valor de valência e um valor de alerta conforme mostra a Tabela 1. Tal base textual é baseada em (Bradley, 1999).

Tabela 1. Exemplo da base textual do Anew-br.
Fonte: Adaptado de Kristensen, 2011

Palavra	Valência		Alerta		Palavra	Valência		Alerta	
	Média	DP	Média	DP		Média	DP	Média	DP
ABALADO	2,58	1,81	5,11	2,82	ARTE	8,08	1,50	3,68	3,21
ABANDONADO	1,85	1,70	5,48	2,49	ÁRVORE	7,98	1,46	2,27	2,19
ABDUÇÃO	5,00	1,85	3,36	2,14	ÁS	5,71	2,00	3,84	2,60
ABELHAS	4,58	2,45	4,38	2,83	ÁSPERO	3,38	2,34	5,49	2,53
ABENÇOADO	8,03	1,84	3,19	3,03	ASSALTANTE	2,03	2,25	7,24	2,70
ABORTO	1,67	1,37	6,42	2,93	ASSALTO	1,47	1,10	7,67	2,25
ABRAÇAR	8,63	1,20	4,30	3,34	ASSAR	6,24	2,73	3,92	2,70
ABRAÇO	8,77	0,75	3,38	3,00	ASSASINO	1,16	0,80	7,51	2,50
ABRASADOR	5,43	1,91	3,53	2,49	ASSENTO	5,87	2,07	3,05	2,20
ABRIGADO	7,61	1,98	3,21	2,41	ASSOVIO	6,15	2,15	4,48	2,62

Segundo Kristensen (2011) a emoção pode ser definida como uma reação temporal breve de prontidão para uma determinada ação, sendo composta por ao menos duas dimensões ortogonais, uma de valência (do desagradável ao agradável) e outra de alerta (do relaxado ao estimulado), em uma concepção definida como a teoria dimensional da emoção.

Para o autor, estímulos que evocam uma

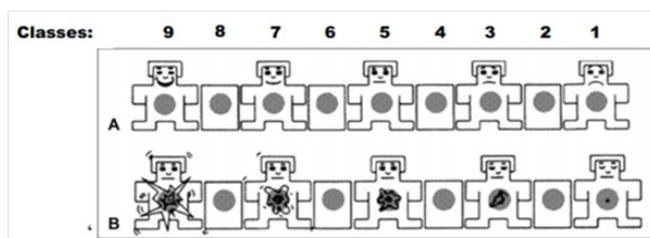
emoção de tristeza, por exemplo, podem ser classificados como de valência desagradável e alerta baixo, enquanto estímulos que evocam uma emoção discreta de felicidade podem ser classificados como de valência agradável e alerta alto; estímulos de raiva podem ser classificados como valência desagradável e alerta alto; etc.

A base textual do Anew-br possui palavras com valores de valência e alerta compreendidas no intervalo de 1 a 9. Sendo assim, palavras que possuem valores baixos, próximos de 1, por exemplo, apresentam valência e alerta baixos, ou seja, desagradável e relaxado, respectivamente. Já palavras com valores próximos de 9 apresentam valência e alerta altos, ou seja, agradável e estimulado, respectivamente.

Nesse trabalho nós agrupamos as palavras do Anew-br em classes da seguinte forma: palavras com valores de valência compreendidas entre 1 e 1,99 pertencem a classe 1, entre 2 e 2,99 pertencem a classe 2, entre 3 e 3,99 a classe 3 e assim sucessivamente. O mesmo foi feito para o alerta. A Figura 2 ilustra esse processo.

Figura 2 - Classes de Valência (A) e Alerta (B).

Fonte: Adaptado de Kristensen, 2011



Como pode ser observado na Figura 2 as

classes 1, 3, 5, 7 e 9 possuem a imagem de um indivíduo que tenta representar as características físicas-sentimentais dessas classes, pois estudos correlatos psicofisiológicos indicam que julgamentos de alerta estão associados a variações na condutância elétrica da pele e julgamentos de valência estão associados a variações na contração de músculos faciais (Kristensen, 2011). As demais classes são consideradas classes intermediárias. A partir disso, foram desenvolvidos modelos matemáticos que emergem um modelo de reputação capaz de estimar a média ponderada de valência (MV) e alerta (MA), conforme as equações 1 e 2:

$$(1) \quad M_V = \frac{q_1 \times V_1 + \dots + q_n \times V_n}{q_1 + \dots + q_n}$$

$$(2) \quad M_A = \frac{q_1 \times A_1 + \dots + q_n \times A_n}{q_1 + \dots + q_n}$$

no qual: MV é a média ponderada para a valência; MA a média ponderada para o alerta; q_i , para $i = 1, \dots, n$, a quantidade de vezes que uma palavra é encontrada; V_i , para $i=1, \dots, n$, o valor de valência de uma palavra; e A_i , para $i=1, \dots, n$, o respectivo valor de alerta de uma palavra.

As equações 1 e 2 foram implementadas na linguagem C, comparando as palavras do arquivo de saída da Figura 1 com a base textual definida no Anew-br, minerando as palavras afetivas em comum. O retorno do algoritmo é a média ponderada para a valência e para o alerta. No pseudocódigo 1 é apresentado o algoritmo para gerar o modelo de reputação proposto.

ENTRADAS:
 ESTRUTURA
 Palavra: STRING;
 Valencia, Alerta: REAIS;
 FIM DA ESTRUTURA;
 SomaValencia, SomaAlerta, M_v , M_a : REAIS;
 V_n : REAL; //Valor de valência de uma palavra n
 A_n : REAL; //Valor de alerta de uma palavra n
 q_n : INTEIRO; //Quantidade de palavras em comum entre o Anew-br e o
 arquivo do facebook
 Arq_Facebook: ARQUIVO; //Arquivo com palavras de um individuo obtidos do facebook
 Arq_Anew: ESTRUTURA; //Base textual do Anew-br

```

01 INICIO:
02 SomaValencia <- 0; SomaAlerta <- 0; qn <- 0;

03 Transfira os dados de Arq_Facebook e Arq_Anew para a RAM;
04 Leia a primeira palavra de Arq_Facebook;
05 ENQUANTO (Não EOF de Arq_Facebook)FAÇA//Enquanto não for final
    de arquivo
06     SE (token) ENTÃO
07         Elimine a primeira e última palavra;
08         Selecione a próxima palavra lida e consulte essa
            palavra em Arq_Anew;
09         SE (Encontrou) ENTÃO
10             SomaValencia <- SomaValencia + Vn;
11             SomaAlerta <- SomaAlerta + An;
12             qn <- qn + 1;
13             Leia a próxima palavra de Arq_Facebook;
14         FIM SE
15     SENÃO
16         Leia a próxima palavra de Arq_Facebook;
17     FIM SENÃO
18 FIM SE
19 SENÃO
20     Selecione a palavra e consulte a mesma em Arq_Anew;
21     SE (Encontrou) ENTÃO
22         SomaValencia <- SomaValencia + Vn;
23         SomaAlerta <- SomaAlerta + An;
24         qn <- qn + 1;
25         Leia a próxima palavra de Arq_Facebook;
26     FIM SE
27     SENÃO
28         Leia a próxima palavra de Arq_Facebook;
29     FIM SENÃO
30 FIM SENÃO
31 FIM ENQUANTO
32 SE (qn <> 0) ENTÃO
33     Mv <- SomaValencia/qn; //Conforme a equação 1
34     Ma <- SomaAlerta/qn; //Conforme a equação 2
35 FIM SE
36 SENÃO
37     Escreva: "Impossível calcular as medias de valência e Alerta";
38 FIM SENÃO
    
```

Pseudocódigo 1. Algoritmo para o modelo de reputação proposto.

Dessa forma foi possível identificar as classes de valência e alerta de um determinado indivíduo, usando como base a representação gráfica

das classes definidas na Figura 2.

RESULTADOS E DISCUSSÃO

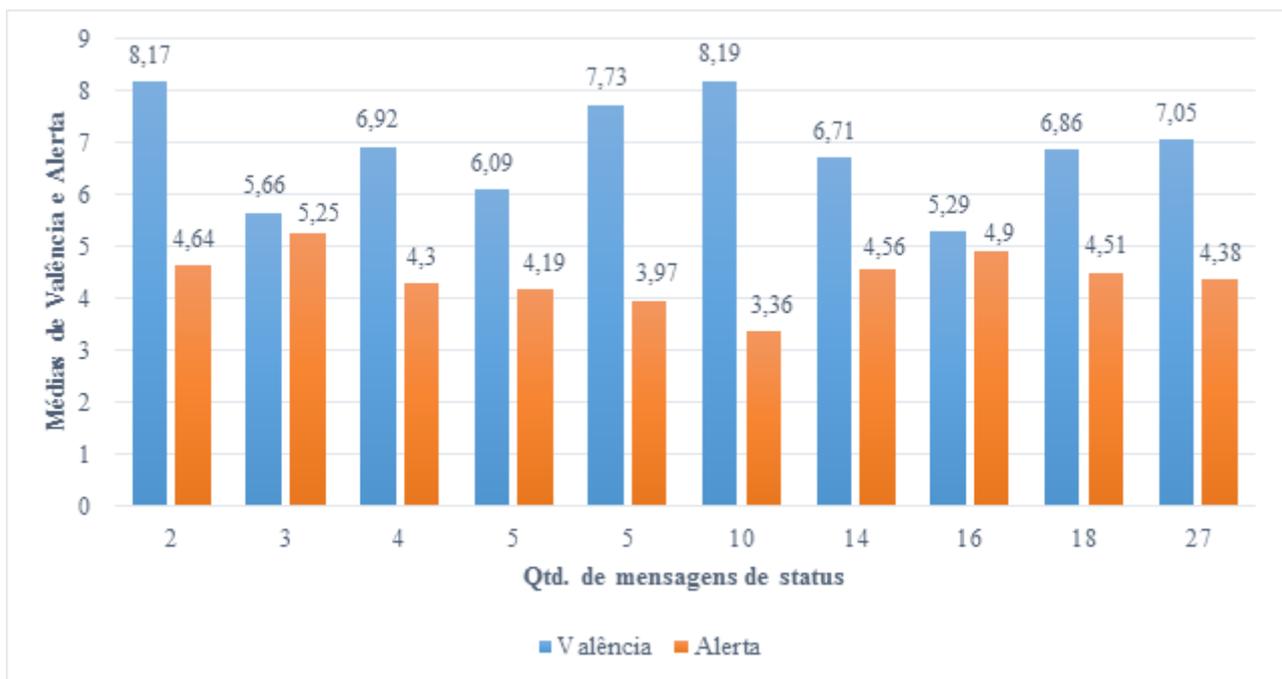
Foram extraídos do Facebook dados de dez indivíduos de nacionalidade brasileira, com idades entre 19 e 23 anos, nascidos nas regiões sul, sudeste e centro-oeste do Brasil. A Tabela 2 e a Figura 3 apresentam os resultados em detalhes.

Tabela 2 – Resultados (Médias de valência e alerta)

Usuários	Sexo	Qtde de mens. status	Média de Valência	Média de Alerta
Indivíduo 1	Feminino	5	6,09	4,19
Indivíduo 2	Masculino	10	8,19	3,36
Indivíduo 3	Feminino	2	8,17	4,64
Indivíduo 4	Feminino	18	6,86	4,51
Indivíduo 5	Feminino	5	7,73	3,97
Indivíduo 6	Feminino	27	7,05	4,38
Indivíduo 7	Masculino	3	5,66	5,25
Indivíduo 8	Masculino	16	5,29	4,90
Indivíduo 9	Feminino	4	6,92	4,30
Indivíduo 10	Masculino	14	6,71	4,56

Figura 3 – Resultados (Médias de valência e alerta)

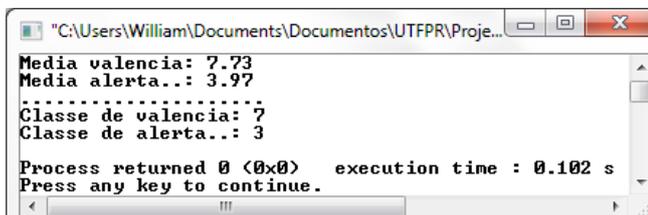
Fonte: Autoria própria, 2014



Conforme a Tabela 2, para cada indivíduo foram calculadas as médias de valência e alerta aplicando o algoritmo do pseudocódigo 1. Para esse teste foram consideradas todas as mensagens de status publicadas por um indivíduo durante um intervalo de uma semana.

Nos testes realizados observou-se que o modelo de reputação proposto é capaz de mostrar a dimensão emocional de um indivíduo conforme apresentado na metodologia. Nos casos em que não houve uma classificação, devido a não haver palavras em comum entre os dados extraídos do Facebook e a base de palavras afetivas do Anew-br, os dados do indivíduo não foram considerados no modelo. A Figura 4 ilustra os resultados de uma extração e classificação.

Figura 4 - Janela de apresentação dos resultados (Classificação)



```

"C:\Users\William\Documents\Documentos\UTFPR\Proje...
Media valencia: 7.73
Media alerta.: 3.97
.....
Classe de valencia: 7
Classe de alerta.: 3

Process returned 0 (0x0)   execution time : 0.102 s
Press any key to continue.
  
```

Conforme ilustrado na Figura 4, observa-se que o indivíduo referente a essa análise foi classificado como valência e alerta 7 e 3 respectivamente, ou seja, a maioria das palavras publicadas por ele pertenciam à essas classes. Logo, com base nas informações contidas na Figura 2 é possível observar seu aspecto emocional.

CONCLUSÕES

Este trabalho apresentou as etapas

desenvolvidas em um projeto baseado na mineração de texto. Modelos matemáticos foram propostos e um modelo de reputação foi construído para identificar o nível emocional de um indivíduo usando os conteúdos postados em suas redes sociais.

Essa abordagem tem sido pouco explorada na literatura e na maioria dos casos é utilizada apenas como forma de avaliação de produtos e serviços a partir de opiniões de usuários, como forma de identificar a reputação de um possível indivíduo ou grupo.

Pragmaticamente, o trabalho procurou identificar e comparar as informações compartilhadas pelos usuários com as palavras afetivas definidas pelo Anew-br usando um modelo de reputação capaz de identificar o nível emocional de um indivíduo.

Os resultados iniciais identificam um índice satisfatório de compatibilidade entre as estimativas providas pelo modelo de reputação e o peso que essa estimativa representa na tabela de referência, ou seja, nos parâmetros base relacionados à lista de palavras do Anew-br.

Ainda assim, diversas questões desafiadoras permanecem, estão sendo analisadas, e compõem as perspectivas de trabalhos futuros. Por exemplo, a base afetiva textual está sendo aumentada (enriquecida) conforme palavras associadas são identificadas nas redes sociais.

Um sistema computacional capaz de obter os dados da rede social em tempo real está sendo

desenvolvido para ser associado ao modelo de reputação apresentado.

Usuários em cenários diferentes serão usados como forma de validar e melhorar a abordagem.

O modelo ajustado poderá ser usado em processos de decisão coletiva ou, ainda, como forma de medir a confiabilidade dos usuários em processos avaliativos.

REFERÊNCIAS

- ALT, C.; ASTRACHAN, O.; FORBES, J.; LUCIC, R.; RODGER, S.. Social Networks Generate Interest in Computer Science. SIGCSE'06 Proceedings of the 37th SIGCSE technical symposium on Computer science education, Volume 38 Issue 1, 2006. 438-442.
- BROMLEY, D. B. Reputation, Image and Impression Management. [S.l.]: Wiley, 1993.
- BRADLEY, M. M.; LANG, P. J. Affective Norms for English Words (ANEW): instruction manual and affective ratings. Technical Report C-1. Gainesville: University of Florida; 1999.
- DOSCIATTI, M. D.; MARTINAZZO, B.; PARAISO, E. C.. Identifying Emotions in Short Texts for Brazilian Portuguese. In: IV International Workshop on Web and Text Intelligence, 2012, Curitiba, Brazil. October/2012. p. 1-10
- FAYYAD, U. M.; PIATETSKY-SHAPIRO, G.; SMYTH, P. From Data Mining to Knowledge Discovery in Databases. Artificial Intelligence Magazine, v. 17, n. 3, p. 37-54, 1996.
- FACEBOOK, <https://pt-br.facebook.com/>, acesso em 15/08/2014.
- JOSANG, A., ISMAIL, R., AND BOYD, C. (2007). A survey of trust and reputation systems for online service provision. Decis. Support Syst., 43(2):618–644. 15, 16, 23
- KRISTENSEN, CHRISTIAN. H. Normas brasileiras para o Affective Norms for English Words. Trends Psychiatry Psychother. 2011, 33(3) – 135-146.
- MUI, L. Computational Models of Trust and Reputation: Agents, Evolutionary Games, and Social Networks. MIT, 2002.
- PINYOL, I., SABATER-MIR, J.: Computational trust and reputation models for open multi-agent systems: a review. Artificial Intelligence Review, V. 40, 2013, 1-25
- PENTAHO, <http://www.pentaho.com/explore/pentaho-data-integration/>, acesso em 15/05/2013.
- SABATER, J.; SIERRA, C. Social regret, a reputation model based on social relations. SIGecom Exch., 3:44–56, 2001.
- SANTOS, LEANDRO. M. Protótipo para mineração de opinião em redes sociais: Estudos de casos selecionados usando o Twitter. 2010. 103 f. Monografia (Bacharel em ciência da computação) – Departamento de ciência da computação, Universidade Federal de Lavras, Lavras, 2010.
- WASSERMAN, S.; GALASKIEWICZ, J. Advances in Social Network Analysis: Research in the Social and Behavioral Sciences. Thousand Oaks: Sage Publications, v. 171, 1994

Artigo submetido em: 31.08.2014

Artigo aceito para publicação em: 20.12.2014