



UNIVERSIDADE FEDERAL DO ESTADO DO RIO DE JANEIRO
CENTRO DE CIÊNCIAS EXATAS E TECNOLÓGICAS
PROGRAMA DE PÓS-GRADUAÇÃO EM INFORMÁTICA

CAPTURA DE PERFIS DINÂMICOS DE USUÁRIOS A PARTIR DA ANÁLISE DE MENSAGENS EM
REDES SOCIAIS

Gustavo Fernandes

Orientador

Prof. Dr. Sean Wolfgang Matsui Siqueira

RIO DE JANEIRO, RJ – BRASIL

SETEMBRO DE 2013

CAPTURA DE PERFIS DINÂMICOS DE USUÁRIOS A
PARTIR DA ANÁLISE DE MENSAGENS EM REDES SOCIAIS

Gustavo Fernandes

DISSERTAÇÃO APRESENTADA COMO REQUISITO PARCIAL
PARA OBTENÇÃO DO TÍTULO DE MESTRE PELO PROGRAMA
DE PÓS-GRADUAÇÃO EM INFORMÁTICA DA UNIVERSIDADE
FEDERAL DO ESTADO DO RIO DE JANEIRO (UNIRIO).
APROVADA PELA COMISSÃO EXAMINADORA ABAIXO ASSINADA.

Aprovada por:



Sean Wolfgang Matsui Siqueira, D.Sc. – UNIRIO



Kate Cerqueira Revoredo, D.Sc. – UNIRIO



Jonice de Oliveira Sampaio, D.Sc. – UFRJ

RIO DE JANEIRO, RJ – BRASIL

SETEMBRO DE 2013

À minha família,

Agradecimentos

Pai, mãe, irmã, esposa e filhote, essa vitória é para vocês que são a razão do meu viver e da minha felicidade. Obrigado a todos pela força e compreensão durante toda essa trajetória que muitas vezes me privaram da convivência com vocês que é sempre tão prazerosa, devido aos compromissos que tinha com os estudos.

Um agradecimento especial ao Renanzinho que depois de dizer algumas vezes a frase: "-Papai, brinca comigo?", ouviu como resposta que o papai não podia brincar porque estava estudando e fazendo o "trabalhinho de casa" da minha escola. Filho, quando você estiver maior e estiver lendo esse texto, espero que não pense que estudar é chato porque não deixa o papai brincar contigo e compreenda a importância de se dedicar aos estudos para sempre evoluir como ser humano e cidadão. Papai vai brincar o dobro com você a partir de agora. Um abraço de urso e nunca se esqueça que o papai te ama muito!

Um agradecimento mais especial ainda à minha linda esposa Juliana que depois do Renanzinho ouvir a desculpa citada no parágrafo anterior, teve que assumir o papel do papai e brincar com o pequeno. Sei que é sempre um prazer conviver e brincar com nosso pequeno, mas às vezes nos deixa estafados porque o moleque não para! Minha bonequinha, tenho certeza que sem você não seria o que sou e não estaria onde estou. Vamos continuar realizando nossos sonhos. Preciso muito de você. Amo-te!

Ao professor Sean pela orientação e ensinamentos. Competente, exigente, apresentou dicas e ideias inteligentes e valiosas, sempre disposto a ajudar.

Aos meus colegas de turma que viraram amigos: Fernando, Ronaldo e Rodrigo. Sempre ajudando e motivando uns aos outros. Conseguimos chegar lá!

E no final de tudo eu posso dizer que valeu. Valeu muito a pena.

Fernandes, Gustavo. **Captura de perfis dinâmicos de usuários a partir da análise de mensagens em redes sociais**. UNIRIO, 2013. 130 páginas. Dissertação de Mestrado. Departamento de Informática Aplicada, UNIRIO.

RESUMO

A cada ano que passa o número de pessoas que acessam a *web* vem crescendo consideravelmente e grande parte desses acessos é originária de sistemas de comunicação e redes sociais. Nestes sistemas, o objeto central é o usuário, pois através dele é possível aprimorar suas funcionalidades para melhor atendê-lo e também melhorar a experiência deste usuário dentro do sistema. É possível obter informações do usuário quando o mesmo se cadastra no sistema, porém estes dados raramente sofrem atualizações. Esta pesquisa apresenta uma forma de conhecer melhor estes usuários mapeando dinamicamente seus perfis através da explicitação de interesses e preferências tendo como insumo mensagens textuais que são inseridas espontaneamente em redes sociais. Após realização de pesquisa qualitativa através de entrevistas on-line com alguns usuários que tiveram seus perfis mapeados na rede social Twitter, se observou que a abordagem teve boa aceitação entre estes usuários e os mesmos acreditam que o perfil construído conseguiu representá-los corretamente.

Palavras-chave: Perfil de usuário, Modelagem de usuários, Redes sociais, Entidade Nomeada, Processamento de Linguagem Natural

ABSTRACT

Over the years the number of people accessing the web has considerably grown and most of the hits come from communication systems and social networks. In order to help these systems to evolve and provide better functionalities, it's crucial to know better the users. It's possible to extract user's information when he signs up on a system, but this information is rarely updated. This research presents an approach to better understand the users' interests and preferences by dynamically mapping their profiles from text messages that are inserted spontaneously in social networks. After conducting qualitative research through online interviews with some users who had their profiles mapped on Twitter social network, it was noticed that the approach was well accepted among these users and they believe that the acquired profile represented them correctly.

Keywords: User profile, User modeling, Social networks, Named entities, Natural language processing

Índice

Capítulo 1 – Introdução	1
1.1. Motivação	1
1.2. Problema	2
1.3. Questões de pesquisa	4
1.4. Objetivos da pesquisa	4
1.5. Abordagem para análise das questões	5
1.6. Metodologia de pesquisa	7
1.7. Organização da dissertação.....	9
Capítulo 2 – Fundamentação Teórica.....	10
2.1. Modelagem de usuários	10
2.1.1. Dimensões	11
2.1.1.1. Conhecimento, crenças e experiência	11
2.1.1.2. Interesses e preferências.....	11
2.1.1.3. Metas, planos, tarefas e necessidades.....	12
2.1.1.4. Informações demográficas	13
2.1.1.5. Estado emocional	13
2.1.1.6. Contexto	13
2.1.2. Representação das dimensões.....	13
2.1.2.1. Modelagem de usuários por sobreposição	13
2.1.2.2. Modelagem de usuários baseada em palavras-chave	14
2.1.2.3. Modelagem de usuários através de estereótipos	14
2.1.3. Dimensões e representação dos usuários na abordagem proposta	15
2.2. Extração de informação	15

2.2.1.	Técnicas de extração de informação.....	16
2.2.2.	Cálculo de pesos	18
2.3.	Trabalhos relacionados	20
Capítulo 3 –	Abordagem Proposta	22
3.1.	Diagrama conceitual	22
3.2.	Arquitetura lógica	24
3.2.1.	Extração de elementos dos <i>posts</i>	26
3.2.1.1.	Montagem do perfil texto.....	26
3.2.1.2.	Montagem do perfil <i>link</i>	27
3.2.1.3.	Montagem do perfil <i>trending topic</i>	28
3.2.2.	Extração de entidades nomeadas e tópicos.....	30
3.2.3.	Montagem dos perfis dinâmicos.....	31
3.3.	Arquitetura física	32
3.3.1.	Extração de elementos dos <i>posts</i>	34
3.3.2.	Extração de entidades nomeadas e tópicos.....	36
3.3.3.	Montagem dos perfis dinâmicos.....	38
Capítulo 4 –	Estudo de Caso.....	40
4.1.	Metodologia	40
4.2.	Preparação do Ambiente	46
4.3.	Realização do Estudo de Caso	47
4.3.1.	Protótipo	48
4.3.2.	Perfil dos Participantes	52
Capítulo 5 –	Avaliação	53
5.1.	Análise inter-participantes e intra-participantes	54
5.1.1.	Como aprimorar a abordagem e a consulta das informações	55

5.1.2. Montagem do perfil dinâmico	57
5.1.3. Extensão da abordagem para outras redes sociais	59
5.2. Análise dos objetivos traçados.....	61
5.3. Ponto de saturação	64
Capítulo 6 – Conclusão	65
6.1. Conclusões	65
6.2. Contribuições	67
6.3. Limitações e generalizações	68
6.4. Trabalhos futuros	69
Referências	72
Anexo I – Roteiro da Entrevista	77
Anexo II – Texto das Entrevistas	80

Índice de Figuras

Figura 1 – Organização da dissertação	9
Figura 2 – Diagrama Conceitual.....	24
Figura 3 – Arquitetura lógica	25
Figura 4 – Exemplo de montagem do perfil texto	26
Figura 5 – Página contendo o conteúdo que será considerado no processamento	27
Figura 6 – Exemplo de montagem do perfil <i>link</i>	28
Figura 7 – Resultado da busca em uma rede social pela <i>hashtag</i> #redesocial	29
Figura 8 – Exemplo de montagem do perfil <i>trending topic</i>	30
Figura 9 – Arquitetura física.....	33
Figura 10 – Exemplo de dados representados em formato JSON	35
Figura 11 – Tela inicial contendo breve explicação sobre a montagem dos perfis	49
Figura 12 – Tela de consulta de informações mostrando as entidades nomeadas do tipo “pessoas”, no 2º quadrimestre de 2012 e no perfil TEXTO	50
Figura 13 – Tela que exhibe os <i>posts</i> relacionados à entidade nomeada “JEFF BEZOS”, no 3º quadrimestre de 2012 e no perfil TEXTO	50
Figura 14 – <i>Post</i> sendo exibido no Twitter	51
Figura 15 – Exibição do texto extraído de um link	51
Figura 16 – Informações sobre os participantes das entrevistas.....	52
Figura 17 – Tipos de perfis apontados como corretos pelos participantes	62
Figura 18 – Ordenação pela representatividade dos tipos de perfis	63

Índice de Tabelas

Tabela 1 - Comparação das características presentes na abordagem proposta	21
Tabela 2 - Diferenças entre pesquisa exploratória e conclusiva (PARASURAMAN, 1991).....	42

Lista de Abreviaturas

API	<i>Application Programming Interface</i>
DPM	<i>Dynamic Profile Mapper</i>
EI	Extração de Informações
HTML	<i>Hyper Text Markup Language</i>
IDF	<i>Inverse Document Frequency</i>
JSON	<i>JavaScript Object Notation</i>
LOD	<i>Linked Open Data</i>
MEDS	Método de Explicitação do Discurso Subjacente
PLN	Processamento de Linguagem Natural
RENC	Reconhecimento de Entidades Nomeadas e Classificação
REST	<i>Representational State Transfer</i>
SQL	<i>Structured Query Language</i>
TF	<i>Term Frequency</i>
TF-IDF	<i>Term Frequency - Inverse Document Frequency</i>
URL	<i>Uniform Resource Locator</i>
WA	Web Adaptativa
WS	Web Semântica

Capítulo 1 – Introdução

Este capítulo fornece uma visão geral da pesquisa, apresentando sua motivação, o problema abordado, as questões de pesquisa, os objetivos a serem alcançados, um resumo da abordagem para análise das questões, a metodologia empregada e por último como o trabalho foi organizado.

1.1. Motivação

O mundo está cada vez mais conectado e a cada dia mais pessoas estão acessando a *web*. Uma pesquisa realizada no final de 2012 mostra que existem mais de dois bilhões de pessoas conectadas e que ainda 2/3 da população mundial ainda não possui acesso à *web* (BUSINESS INSIDER, 2012).

Grande parte desses acessos provém de sistemas de comunicação ou redes sociais, como, por exemplo, o Facebook¹. Para se ter uma ideia da importância das redes sociais na vida cotidiana, dados mostram que 1/7 da população mundial possui perfil criado no Facebook e que no final do ano de 2011, os norte-americanos passaram mais tempo acessando redes sociais (em torno de 26 minutos por dia) do que os tradicionais portais de notícias que são acessados em média 18 minutos por dia (BUSINESS INSIDER, 2012).

¹ Site: <http://www.facebook.com>

Nas redes sociais, o objeto central é o usuário, pois através dele é possível aprimorar as funcionalidades e também melhorar sua experiência dentro do sistema. Em função disso, é muito importante que se conheça cada vez mais o perfil dos usuários, mapeando, por exemplo, seus interesses e preferências para que esses sistemas possam, na medida do possível, se adequar cada vez melhor a esses usuários. Essa adequação serve não só para as redes sociais proverem uma experiência de navegação mais agradável, como também para a exploração comercial através da exibição de anúncios e ofertas que se enquadrem em seus perfis. Para isso não é viável se concentrar apenas nos atributos que são informados inicialmente no momento que os usuários se cadastram nessas redes sociais, pois além de muitos deles não terem obrigatoriedade de preenchimento, estes usuários não se preocupam em ficar periodicamente atualizando essas informações fornecidas inicialmente. Por isso torna-se muito importante conhecer o perfil desses usuários baseando-se nas informações que os mesmos inserem de forma espontânea a cada dia nestas redes sociais.

1.2 Problema

Dentre as direções mais promissoras de desenvolvimento da *web*, dois paradigmas existentes compartilham o objetivo de proporcionar aos usuários um acesso mais eficiente às informações e serviços on-line: a *Web Adaptativa* (WA) (PERKOWITZ & ETZIONI, 1997) e a *Web Semântica* (WS) (BERNERS-LEE *et al.*, 2001). O primeiro tem como objetivo permitir experiências personalizadas de disponibilização de conteúdos (SOSNOVSKY & DICHEVA, 2010). O segundo propõe um conjunto de padrões e tecnologias para prover uma descrição significativa (enriquecida semanticamente) e permitir o processamento automático do conteúdo da *web* (estruturado em forma de dados) e transferência de conhecimento na *web*. Os desafios

da WA introduzem áreas de aplicação interessantes para as tecnologias da WS, enquanto a visão da WS, por sua vez, abre novas perspectivas para a pesquisa da WA (SOSNOVSKY & DICHEVA, 2010).

Para que ocorra a adaptação na *web*, ou seja, um determinado serviço ou portal se adapte ao seu público, é necessária a modelagem de usuários, que é uma área de pesquisa mais antiga que WA. A modelagem de usuários tem por objetivo delinear as características dos usuários e os primeiros sistemas que modelavam usuários foram desenvolvidos em torno de 1970 (CARBONELL, 1970). A precisão das premissas assumidas na modelagem dos usuários define a eficácia dos sistemas adaptativos que são sistemas que se adaptam às características dos usuários (SOSNOVSKY & DICHEVA, 2010).

De forma geral, quando um usuário da *web* se cadastra em alguma rede social é questionado com uma série de informações que nem sempre são necessárias para utilização desse sistema. Além disto, com o passar do tempo, estes usuários não costumam atualizar essas informações inseridas inicialmente a não ser que os mesmos sejam obrigados a fazê-lo. Na área de modelagem de usuários, esse tipo de perfil criado inicialmente que dificilmente sofre atualizações é chamado de perfil estático (HOTHY & HALL, 1998).

Depois de algum tempo utilizando redes sociais, estes usuários inserem várias mensagens e essas informações muitas vezes não são consideradas para capturar dados importantes dos usuários fazendo com que as redes sociais utilizadas não consigam acompanhar a evolução dos perfis destes usuários.

É justamente esse tipo de situação que este trabalho se propõe a explorar, ou seja, capturar as informações inseridas espontaneamente pelos usuários em redes sociais para montar seu perfil de forma automática.

1.3 Questões de pesquisa

Considerando o problema proposto, duas questões de pesquisa são formuladas e exploradas ao longo do trabalho. São elas:

- Utilizando-se das mensagens inseridas por usuários em redes sociais e dos elementos presentes nas mesmas, é possível construir o perfil desses usuários de forma dinâmica que sejam representativos de seus interesses?
- Processando-se os textos de cada elemento presente nas mensagens dos usuários e construindo um tipo distinto de perfil para cada elemento, qual representará melhor o perfil dos usuários?

Os elementos presentes nas mensagens inseridas pelos usuários citadas nas questões da pesquisa são: (i) o próprio texto das mensagens, (ii) os *links* que são inseridos junto aos textos e (iii) marcações específicas que servem para categorizar e agrupar mensagens que se referem a um determinado tópico e são chamadas comumente de *hashtags*. As *hashtags* mais populares, ou seja, as que mais são referenciadas nos *posts* dos usuários compõem os *trending topics*, que podem ser entendidos como os assuntos que estão em foco no momento. As *hashtags* iniciam-se com o caractere “#” e alguns exemplos são *#usopen* (representando um torneio de tênis norte-americano) e *#copadobrasil* (representando um torneio de futebol que ocorre no Brasil).

Estes três elementos são comumente encontrados em algumas redes sociais amplamente utilizadas, como Twitter², Facebook e Google Plus³.

1.4 Objetivos da pesquisa

Este trabalho propõe a captura do perfil de usuários de redes sociais de forma

² Site: <http://www.twitter.com>

³ Site: <http://plus.google.com>

dinâmica baseando-se no processamento e análise das mensagens textuais inseridas por estes usuários, ou seja, construir o perfil do usuário dinamicamente desconsiderando os atributos inseridos inicialmente na rede social (perfil estático) que não sofrem mudanças durante o tempo.

Foi construída uma arquitetura chamada *Dynamic Profile Mapper* (DPM) e a mesma consiste no desmembramento de cada mensagem textual inserida pelos usuários em três componentes principais: o texto da mensagem, os *links* inseridos e as *hashtags* inseridas.

Baseado nestes três elementos, o DPM constrói três perfis distintos para verificar se algum(ns) dele(s) representou(aram) corretamente o perfil dos usuários e realiza algumas análises e comparações entre os três.

Para responder as questões de pesquisa, um objetivo principal e um secundário foram definidos. São eles:

- **Objetivo Principal** - Capturar dinamicamente o perfil dos usuários verificando se os mesmos acreditam que foi gerado um perfil representativo deles na rede social objeto do estudo.

- **Objetivo Secundário** - Verificar entre os perfis existentes, qual foi o mais representativo dos interesses dos usuários.

Para atingir os dois objetivos, foi desenvolvido um protótipo que coletou e processou mensagens inseridas por um grupo de usuários na rede social Twitter durante o ano de 2012 e após processar os dados, uma pesquisa qualitativa foi conduzida junto a esses usuários. Os dados coletados foram analisados e serão apresentados no capítulo 5.

1.5 Abordagem para análise das questões

A solução proposta concentra-se primordialmente na busca de informações que são inseridas espontaneamente por usuários em redes sociais. As informações em

questão são as mensagens textuais que estes usuários cadastram corriqueiramente, mensagens estas chamadas de *posts*.

Considerando estes *posts* e os três elementos citados anteriormente (texto das mensagens, *links* e *hashtags*), foi possível construir um tipo de perfil de usuário distinto para cada elemento. São eles: perfil texto, perfil *link* e perfil *trending topic*. Estes perfis são compostos por um conjunto de entidades nomeadas extraídas dos textos que representam os interesses e preferências dos usuários. Além disso, cada mensagem textual inserida também é categorizada por um assunto e esses assuntos são consolidados de forma a explicitar para cada usuário quais são os assuntos de seu maior interesse.

O termo “entidade nomeada” amplamente utilizado em processamento de linguagem natural⁴ foi apresentado por GRISHMAN & SUNDHEIM (1996), buscando-se reconhecer unidades de informação, como nomes, incluindo pessoas, organizações e locais, além de expressões numéricas, incluindo tempo, datas, moedas e expressões percentuais. Identificar referências a estas entidades nos textos foi reconhecido como uma sub-tarefa importante de extração de informações (EI)⁵ e foi chamado de "reconhecimento de entidades nomeadas e classificação (RENC)" (NADEAU & SEKINE, 2007).

Estas entidades são categorizadas da seguinte forma: organizações, pessoas, lugares, eventos e outros, sendo esta última categoria uma agregação de entidades que não se encaixam nas categorias anteriores.

SOSNOVSKY & DICHEVA (2010) citam que interesses e preferências representam uma das dimensões existentes para modelar de usuários, dentre outras como informações

⁴ Processamento de linguagem natural (PLN) é uma subárea da inteligência artificial e da linguística que utiliza um conjunto de métodos formais para analisar textos e estudar os problemas da geração e compreensão automática de línguas humanas naturais. Tem como objetivo fornecer aos computadores a capacidade de reconhecer o contexto, fazer análise sintática, semântica, léxica, morfológica e extrair informação dos textos processados.

⁵ GAIZAUSKAS & WILKS (1998) definem extração de informação como a atividade de extrair automaticamente uma informação de um texto em língua natural. A informação extraída é determinada por um conjunto de padrões ou regras de extração específicas ao seu domínio: padrões que podem ser escolhidos de maneira manual, por algum especialista, ou de forma automatizada.

demográficas ou estado emocional. Para representar estas dimensões, SOSNOVSKY & DICHEVA (2010) enumeram algumas representações válidas e uma delas é modelar o perfil dos usuários por palavras-chave, aqui representadas por entidades nomeadas. Outros exemplos de modelagem seriam modelagem de usuários por sobreposição e modelagem de usuários através de estereótipos, que serão explicados no Capítulo 2.

1.6 Metodologia de pesquisa

Inicialmente foi elaborada uma revisão bibliográfica ampla, com a finalidade de analisar tudo o que estava sendo publicado em relação à modelagem de usuários, extração de informação e processamento de linguagem natural. Em seguida, foi feita uma filtragem e seleção dos artigos, dissertações, teses e livros que tratam desses tópicos. Todo esse material foi organizado e catalogado no *software* Mendeley⁶ para que fosse possível encontrá-los facilmente, pois este *software* disponibiliza consultas por palavras-chave.

Para responder as perguntas da pesquisa foi necessário o desenvolvimento de um protótipo que apoiou a execução de um estudo de caso através de uma pesquisa qualitativa realizada com usuários que fazem uso de redes sociais.

Esta pesquisa tem um caráter exploratório-descritivo e qualitativo. É importante também ressaltar que as pesquisas exploratórias são aquelas que têm por objetivo explicitar e proporcionar maior entendimento de um determinado problema. Nesse tipo de pesquisa, o pesquisador procura um maior conhecimento sobre o tema em estudo (GIL, 2005).

O estudo de caso foi escolhido como o método da pesquisa, pois esta é uma pesquisa empírica que foi realizada em um contexto real. Estudo de caso deve ser usado

⁶ Site: <http://www.mendeley.com>

quando o pesquisador tem pouco controle sobre os acontecimentos, em contextos reais e sem controle total sobre as variáveis (YIN, 2005).

A pesquisa qualitativa foi realizada através de entrevistas on-line elaboradas seguindo o Método de Explicitação do Discurso Subjacente (MEDS) (NICOLACI-DA-COSTA, 2007), que é um método exploratório o qual tem em comum com os demais métodos qualitativos, a investigação aprofundada em contexto, as amostras pequenas e a flexibilidade de procedimentos e técnicas. A principal diferença entre o MEDS e grande parte dos demais métodos qualitativos é o registro explícito de que trabalha com material discursivo, aliado a uma concepção de que o discurso desempenha um importante papel na construção social da esfera do psicológico (NICOLACI-DA-COSTA, 2007).

Segundo o MEDS, inicialmente é necessário delinear o(s) objetivo(s) da pesquisa e com os objetivos definidos, um roteiro de entrevista foi elaborado contendo 13 perguntas (em sua maioria abertas) e entrevistas on-line foram executadas com todos os participantes da pesquisa.

Em posse dos dados das entrevistas, seguiu-se para o próximo passo chamado pelo MEDS de análise inter-participantes que tem por objetivo fornecer ao pesquisador uma visão geral dos resultados obtidos através da análise de todos os depoimentos coletados, buscando-se encontrar estruturas regulares, padrões e elementos recorrentes nas entrevistas. Em conjunto com esta análise foi realizada a análise detalhada de cada uma das entrevistas individuais, denominada pelo MEDS de análise intra-participantes. Nesta fase é realizada uma análise sistemática das respostas individuais, na qual são feitas comparações internas aos depoimentos de cada um dos entrevistados. Estas comparações buscam inconsistências, contradições, novos conceitos, novos usos de linguagem, etc. no discurso de cada um dos participantes (NICOLACI-DA-COSTA *et al.*, 2009).

1.7 Organização da dissertação

Um diagrama contendo os principais tópicos desta pesquisa é apresentado na Figura 1. Estes tópicos foram mapeados em todos os capítulos deste trabalho através das setas tracejadas.

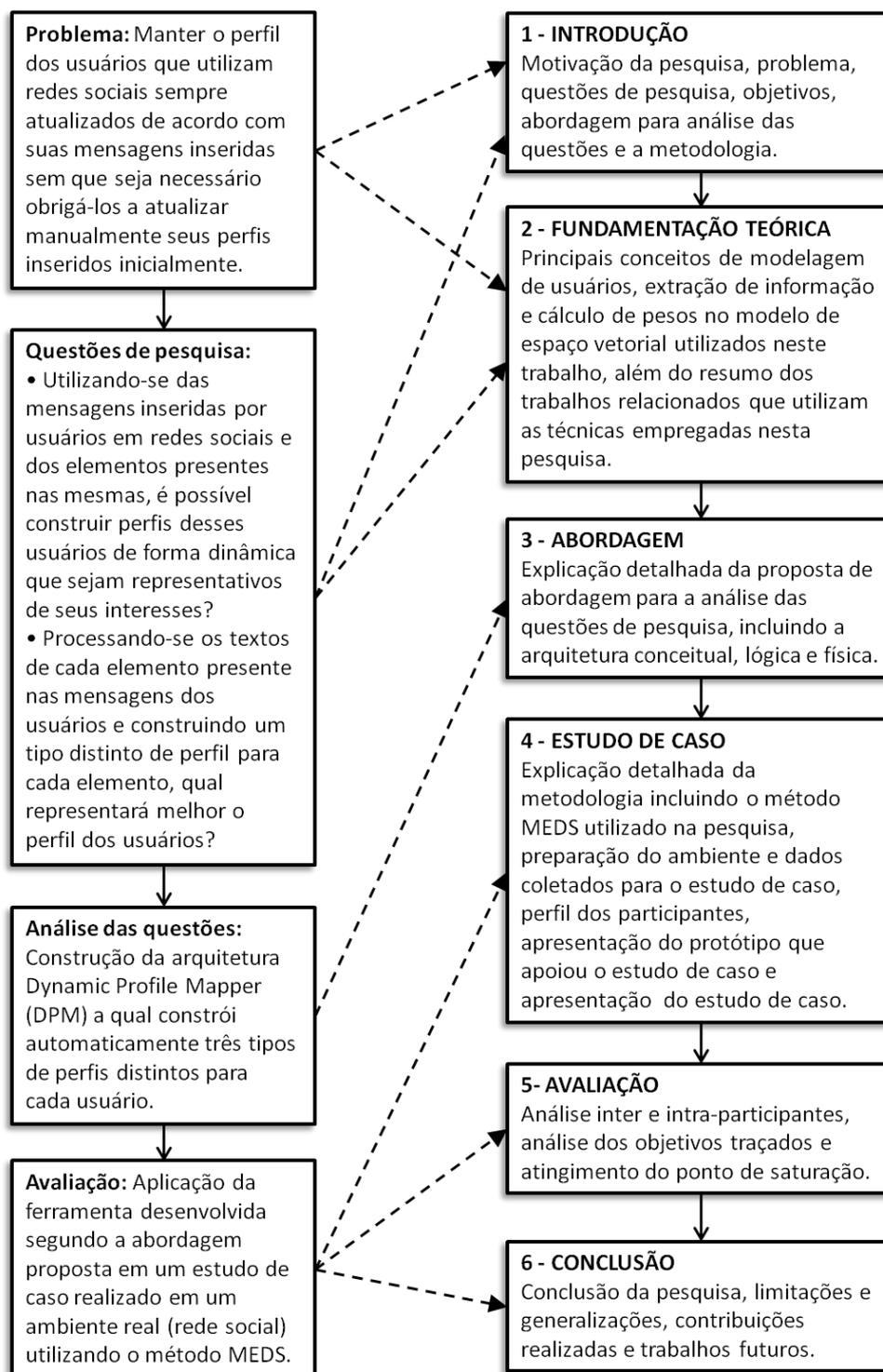


Figura 1 – Organização da dissertação

Capítulo 2 – Fundamentação Teórica

Neste capítulo serão apresentados alguns conceitos fundamentais para o entendimento deste trabalho. Os conceitos relacionados à modelagem de usuários são necessários para entender como os usuários podem ser representados dentro dos mais variados tipos de sistemas. Os conceitos de extração de informação serão necessários para entender as atividades aplicadas na solução proposta para extrair as entidades nomeadas dos textos presentes nos *posts* inseridos pelos usuários, explicitando, inclusive, o cálculo de pesos utilizado para estabelecer critérios de seleção das entidades nomeadas para selecionar quais entidades foram as mais representativas para cada tipo de perfil construído. Por último, são apresentados alguns trabalhos relacionados que utilizam as técnicas empregadas neste trabalho.

2.1. Modelagem de usuários

O objetivo da análise e modelagem de usuários é identificar quem são os usuários e caracterizá-los, isto é, especificar quais funções exercem, quais capacidades possuem, etc. (de SOUZA *et al.*, 1999). Sistemas que necessitam personalizar suas funcionalidades de acordo com o modelo do usuário precisam informar os elementos certos, no momento certo e da forma certa para este usuário (FISCHER, 2001) e para fazê-lo é necessária uma representação interna do usuário.

A ideia central de adaptação de um sistema, isto é, fazer com que o mesmo se adapte aos usuários é baseada na suposição de que diferenças em algumas características do usuário afetam a utilidade dos serviços ou informações fornecidos para estes usuários (SOSNOVSKY & DICHEVA, 2010). Assim, se o comportamento de um sistema é adaptado de acordo com tais características, o seu valor para os usuários aumentará.

2.1.1. Dimensões

Segundo SOSNOVSKY & DICHEVA (2010), é possível mapear algumas dimensões no que diz respeito à adaptação. Estas dimensões são listadas a seguir:

2.1.1.1. Conhecimento, crenças e experiência

Estas características são especialmente importantes para modelagem de alunos em sistemas adaptativos educacionais. Para estes tipos de sistemas, o conhecimento do aluno é a característica mais importante a ser considerada, pois é através dele que estes sistemas definem a forma correta de adaptação para estes alunos. Alguns autores argumentam que o modelo do usuário, no caso o aluno, nesses sistemas adaptativos voltados para a educação, não deve diferenciar entre o conhecimento correto ou incorreto de um conceito e sim tratá-los como crenças deste aluno (SELF, 1988). Por último, experiência é qualquer vivência relevante que o usuário adquiriu antes da utilização do sistema em questão.

2.1.1.2. Interesses e preferências

São muito utilizadas no contexto da WA que consideram estas características as

mais importantes para sistemas desenvolvidos nesta área. Quando se modelam interesses e preferências dos usuários, os sistemas tipicamente são distinguidos entre modelos de longo prazo e modelos de curto prazo. Interesses de longo prazo são relativamente estáveis e evoluem lentamente ao longo de todo o período de interação do usuário com o sistema, ou são fornecidos explicitamente pelo usuário através do preenchimento de informações organizadas em categorias gerais. Um modelo de interesses de curto prazo é dinâmico e, geralmente, preenchido e utilizado durante uma única sessão. O modelo de curto prazo é descartado quando a sessão de uso do sistema é finalizada.

2.1.1.3. Metas, planos, tarefas e necessidades

Modelagem de metas e planos dos usuários tem sido amplamente explorada em sistemas de diálogo inteligentes que se baseiam em processamento de linguagem natural, onde os mesmos compreendem, analisam e tiram conclusões a partir das informações passadas a eles e são usados para a comunicação com humanos. Conhecimento de que situação um usuário está tentando atingir (meta) e que sequência de ações o levará para um estado desejado (plano) é essencial para tais sistemas com o objetivo de manter uma conversa adaptativa com o usuário (KASS & FININ, 1988). Muito próximo a essas dimensões de modelagem, encontra-se o conceito de tarefa ou necessidade. Modelar necessidades dos usuários e buscar ou mapear tarefas é uma abordagem muito popular empregada por diferentes sistemas adaptativos na web, como, por exemplo, sistemas adaptativos de recomendação (McNEE *et al.*, 2006).

2.1.1.4. Informações demográficas

Para alguns tipos de sistemas adaptativos como sistemas de e-commerce, é vital estar ciente das informações demográficas de um usuário desde as características mais básicas, como sexo, idade ou a língua nativa, até as mais complexas, como parâmetros sócio-culturais, nível de educação formal e renda familiar.

2.1.1.5. Estado emocional

Um sistema adaptativo capaz de modelar as emoções de um determinado usuário obtém uma fonte de informação a mais para adaptar e justificar de forma adequada o comportamento do usuário neste sistema.

2.1.1.6. Contexto

O contexto de um usuário é um conceito muito amplo que pode incluir qualquer informação sobre a localização de um usuário, ambiente físico e social, o dispositivo que está sendo utilizado para acessar o sistema, entre outros.

2.1.2. Representação das dimensões

Baseado nas dimensões listadas anteriormente, (SOSNOVSKY & DICHEVA, 2010) citam três formas principais de como as mesmas podem ser representadas ou modeladas:

2.1.2.1. Modelagem de usuários por sobreposição

Esta abordagem é a mais antiga para a representação do modelo de usuários e foi tradicionalmente utilizada por diferentes tipos de sistemas adaptativos educacionais para

a modelagem do conhecimento do aluno como um subconjunto do conhecimento do especialista de um determinado domínio. Pode ser definida como uma técnica para descrever competências dos usuários como um conjunto de hipóteses sobre sua familiaridade com os elementos de representação de um domínio fornecidos por um especialista (CARR & GOLDSTEIN, 1977). Este tipo de modelagem depende do modelo do domínio tal qual um *template* e, essencialmente, consiste de um conjunto de pares conceito-valor, onde o valor representa uma avaliação de um determinado conceito.

2.1.2.2. Modelagem de usuários baseada em palavras-chave

Teve sua origem nas áreas de recuperação e filtragem de informação (BELKIN & CROFT, 1992), em que o conteúdo de um documento é tradicionalmente representado como um vetor de termos (palavras-chave) extraído do texto. Este tipo de modelagem necessita de tecnologias de processamento de linguagem natural (PLN) para sanar problemas como a homonímia (múltiplos significados de uma palavra) e sinonímia (várias palavras que expressam o mesmo significado).

2.1.2.3. Modelagem de usuários através de estereótipos

O principal objetivo dos sistemas adaptativos é ajustar o seu comportamento às características de cada usuário. No entanto, para algumas tarefas é possível identificar categorias típicas de usuários que utilizam o sistema de uma maneira similar e a partir disto esperar resultados semelhantes com o objetivo de disponibilizar um conjunto de funcionalidades semelhantes para esses usuários. Tais categorias que constituem pontos fortes de semelhanças entre os usuários são chamados de estereótipos. Modelagem de usuários através de estereótipos é vantajoso quando, a partir de uma pequena evidência

sobre um usuário, o sistema infere uma grande quantidade de informações do mesmo quando se encaixa em um estereótipo.

2.1.3. Dimensões e representação dos usuários na abordagem proposta

Neste trabalho as dimensões consideradas no perfil dos usuários foram interesses e preferências e as mesmas foram representadas através da modelagem de usuários baseada em palavras-chave. Isso se justifica pelas características da pesquisa, pois a mesma se baseia em redes sociais onde mensagens textuais inseridas pelos usuários e textos presentes em diversos *links* e *hashtags* são processados, o que indica fortemente a necessidade de processamento de linguagem natural e tanto as dimensões escolhidas, quanto a representação do perfil dos usuários são as que mais se adequam a essa situação.

2.2. Extração de informação

Extração de informação tem por objetivo processar informações não estruturadas para buscar entidades e relações entre as mesmas que possam ser representadas de maneira estruturada (SARAWAGI, 2008). Seu objetivo é agilizar o acesso à informação, visto que essa busca pode ser bastante complexa e demorada se fosse realizada por humanos.

A extração de informação estuda técnicas para a obtenção de informações específicas de um domínio a partir de conteúdos textuais em linguagem natural. As principais características que podem ser encontradas em uma técnica de extração de informação são: (i) necessidade de informação a ser obtida de uma fonte de dados normalmente composta por conteúdo textual ou imagens; (ii) impossibilidade de obter a resposta por humanos devido ao grande volume de dados; (iii) inexistência de estrutura

formal nos dados como a obtida em um banco de dados, por exemplo, o que não permite uma consulta direta feita por computadores (MOENS, 2006).

O reconhecimento de entidades visa extrair e classificar qualquer elemento (conceito) do mundo real, concreta ou abstrata, mencionada no texto. Já a relação entre entidades busca por elos semânticos entre as entidades.

Informações semânticas podem ser extraídas de informações não estruturadas com a aplicação de técnicas que utilizam padrões de extração, manualmente construídos ou “aprendidos” em coleções de documentos (MOENS, 2006). Estes padrões são obtidos a partir da organização linguística do conteúdo das informações, sendo possível utilizar seus elementos léxicos que são palavras e grupos de palavras, construções gramaticais que são frases, sentenças e expressões temporais, além da estrutura discursiva que é composta por parágrafos, capítulos etc. (PRATES, 2011).

Importante ressaltar que extração de informação não é o mesmo que recuperação de informação. Técnicas de recuperação de informação podem localizar documentos pertinentes dentro de uma coleção (*corpus*), mas não conseguem extrair informação destes documentos seguindo critérios específicos. A grande vantagem de um sistema de extração de informação em relação a um sistema de recuperação de informação está na habilidade de conseguir extrair informação relevante dos textos de acordo com uma série de critérios definidos e de representá-los em estruturas, algo que sistemas de recuperação de informação não conseguem produzir (SCARINCI, 2000).

2.2.1. Técnicas de extração de informação

As técnicas de extração de informação podem ser classificadas em duas categorias: técnicas simbólicas e técnicas de reconhecimento de padrão, sendo que as

técnicas de reconhecimento de padrão podem ainda ser divididas em supervisionadas e não supervisionadas.

As técnicas simbólicas são aquelas que utilizam regras para a extração definidas manualmente com a ajuda de um especialista no domínio da aplicação (MOENS, 2006). Em um cenário típico, o especialista tem acesso a uma coleção de documentos (*corpus*) de tamanho moderado que seja representativo do domínio da aplicação. Este *corpus* é inspecionado para a definição das regras de extração (PRATES, 2011).

As técnicas de reconhecimento de padrões, também conhecidas como classificação de padrões, têm por objetivo classificar dados baseados tanto em conhecimentos prévios adquiridos por especialistas quanto em conhecimento automaticamente aprendido a partir de dados coletados. A classificação de padrões pode ser vista como uma combinação de recursos e seus valores (MOENS, 2006). No caso da extração de informação, esses recursos são características textuais que podem ser identificadas ou medidas, como, por exemplo, termos e pesos associados a estes termos, e que possuam valores discriminativos quando da organização de padrões em classes semânticas (PRATES, 2011).

A maioria das aplicações em extração de informação utiliza técnicas supervisionadas de reconhecimento de padrões. Nestas técnicas, o classificador treina a partir de uma coleção de exemplos previamente classificados. A ideia geral é que um especialista possa anotar os fragmentos que deveriam ser extraídos em um pequeno subconjunto de documentos para que então o sistema possa generalizar a partir destes exemplos e produzir uma função ou regra que possa ser aplicada ao resto dos documentos. A aplicabilidade destas técnicas está no fato de ser mais fácil anotar documentos do que escrever regras de extração (PRATES, 2011).

Embora muito utilizadas em extração de informação, as técnicas supervisionadas tem como principal crítica o fato de que funcionam bem somente em domínios fechados (baseados em bases de conhecimento ou *corpus* pré-analisados). Considerando domínios abertos (considera uma vasta coleção de documentos que não são pré-analisados, onde a extração se concentra na aprendizagem de entidades), o maior problema reside na falta de exemplos anotados suficientes para o aprendizado. Já os algoritmos de agrupamento são completamente não supervisionados e não dependem de dados anotados. Estes algoritmos conseguem detectar padrões similares em agrupamentos, os quais permitem a descoberta de similaridades e diferenças entre os padrões e derivam conclusões úteis sobre estes agrupamentos (PRATES, 2011).

2.2.2. Cálculo de pesos

O cálculo de pesos define a importância de todos os termos que ocorrem em consultas ou documentos e são representados através de um valor numérico. Seu uso é muito importante para definir o que deverá ser levado em consideração quando algum texto ou documento é processado automaticamente.

Segundo CUMMINS (2008), a frequência de ocorrências de termos foi proposta pela primeira vez como uma medida da utilidade de um termo por LUHN (1958). Luhn desenvolveu um método de contagem para selecionar palavras significativas em um *corpus*. Sua técnica classificou as palavras em ordem de frequência e empregou dois pontos de corte. Um ponto de corte superior que rejeita os termos com melhor ranqueamento, termos que são muito frequentes para serem utilizados na distinção de documentos e um ponto de corte inferior que rejeita termos com classificações muito baixas, uma vez que os mesmos são muito raros e não contribuem significativamente para o conteúdo do artigo. Assim, a técnica mede o que ele chamou de poder de

resolução e promoveu termos que não eram nem muito raros, nem muito frequentes. Apesar da ideia de Luhn de usar a frequência de ocorrência dos termos ser bastante simples, foi essencialmente o começo de muitas das ideias que são empregados atualmente para métodos de cálculo de pesos.

Existem algumas maneiras de calcular estes pesos, sendo *Term Frequency* (TF) e *Term Frequency - Inverse Document Frequency* (TF-IDF) duas formas de cálculo, as quais compartilham a ideia de considerar a frequência de ocorrência dos termos nos textos que serão processados.

Utiliza-se TF para evidenciar os termos mais relevantes e que melhor representam um determinado documento, considerando como peso exatamente o número de ocorrências (frequência) do termo neste documento e pode ser denotado por $TF_{t,d}$, onde t representa o termo e d o documento.

TF-IDF é utilizado como uma medida estatística para avaliar o quão importante é uma palavra em um documento dentro de uma coleção ou *corpus*. A importância aumenta proporcionalmente com o número de vezes que uma palavra aparece no documento, mas é compensado pela frequência da palavra no *corpus*. TF-IDF é muito útil para diferenciar um termo dentro de uma coleção, isto é, demonstrar sua importância em relação ao *corpus*. Esta medida aumenta se o termo ocorre em um documento e diminui se ocorre no *corpus* com grande frequência, denotando assim sua originalidade e pode ser representado pela seguinte equação: $TF-IDF = TF \times IDF$, sendo $IDF = \log(N/df)$, onde N representa o número total de documentos do *corpus* e df o número de documentos onde o termo aparece.

Para deixar mais claro os conceitos de TF e TF-IDF, observemos a seguinte situação: consideremos um documento que contenha 1.000 palavras, onde a palavra “futebol” apareça 30 vezes. O cálculo TF para “futebol” seria $30/1.000$, ou seja, 0,03.

Agora, suponhamos que existem 100.000.000 de documentos e a mesma palavra “futebol” apareça em 10.000 destes. Então, o cálculo TF-IDF é feito através do log (100.000.000 / 10.000) que é igual a 4 e depois calcula-se o produto destas quantidades (0,03 * 4) que resulta no valor 0,12.

2.3. Trabalhos relacionados

Em seu trabalho, SUGIYAMA *et al.* (2010) apresentam uma maneira de realizar buscas personalizadas na web de acordo com o perfil de cada usuário com base no histórico de navegação desses usuários e de usuários com perfis semelhantes, mas a pesquisa ignora todas as ações realizadas pelos usuários em redes sociais.

Twitterrank, proposto por WENG *et al.* (2010), é uma extensão do algoritmo PageRank, que mede a influência dos usuários no Twitter, considerando a estrutura de *links* e a similaridade entre os tópicos extraídos de mensagens dos usuários. Apesar da análise de informações das mensagens dos usuários para extrair tópicos, o conteúdo de todas as mensagens textuais não foi processado para melhorar a sua proposta. Um trabalho similar foi desenvolvido por MISLOVE *et al.* (2010), os quais propuseram inferir perfis de usuários em uma rede social apenas considerando alguns atributos existentes de uma parte dos usuários e seu algoritmo poderia prever os atributos dos demais usuários daquela rede. Eles construíram uma rotina no algoritmo para recuperar automaticamente informações dos usuários no Facebook, mas sua proposta ignora todo o conteúdo das mensagens textuais inseridas pelos usuários e considera somente alguns atributos estáticos. Talvez se nesta pesquisa fosse considerado e processado o conteúdo das mensagens textuais, a precisão dos perfis inferidos poderia crescer significativamente.

Um *framework* para modelagem de usuários no Twitter que enriquece a semântica das mensagens inseridas é apresentado por ABEL *et al.* (2011). Os dados coletados são utilizados para montar diferentes tipos de perfis de usuários, que são comparados entre si e a avaliação é feita através da análise de uma amostra de dados definida pelos próprios autores da pesquisa. A análise dos dados e comparação do perfil é meramente quantitativa e não considera os dados qualitativos, ou seja, os donos dos perfis não participaram da pesquisa e desta forma os autores inferem que os perfis foram construídos corretamente. Outro aspecto a se considerar é a arquitetura, que foi construída para ser utilizada apenas no Twitter, não permitindo flexibilidade suficiente para usá-la em outras redes sociais.

Na tabela abaixo é feita a comparação das principais características presentes na abordagem proposta entre todos os trabalhos relacionados:

Tabela 1 - Comparação das características presentes na abordagem proposta

	Construção dos perfis dos usuários de forma dinâmica	Análise de mensagens textuais inseridas pelos usuários	Arquitetura flexível para utilização em diversas redes sociais	Categorização por interesses ou tópicos	Análise dos perfis construídos pelos próprios usuários
ABEL et al. (2011) - Analyzing User Modeling on Twitter for Personalized News Recommendations	X	X		X	
MISLOVE et al. (2010) - You are who you know: Inferring user profiles in online social networks	X		X		
SUGIYAMA et al. (2010) - Adaptive Web Search Based on User Profile Constructed without Any Effort from Users	X				X
WENG et al. (2010) - TwitterRank: finding topic-sensitive influential Twitterers	X	X		X	

Capítulo 3 – Abordagem Proposta

Neste capítulo serão apresentados o diagrama conceitual e as arquiteturas lógica e física da abordagem proposta. As arquiteturas lógica e física são divididas em três módulos: extração de elementos dos *posts*, extração de entidades nomeadas e tópicos, e montagem dos perfis dinâmicos.

3.1. Diagrama conceitual

A abordagem proposta monta o perfil dinâmico dos usuários através da análise das mensagens (*posts*) inseridas pelos mesmos em redes sociais. Estas mensagens muitas vezes possuem *links* que apontam para outras páginas *web*, tais como: portais de notícias, blogs etc. e até mesmo para outras redes sociais. Outro elemento que também aparece nestas mensagens são marcações que servem para contextualizá-las, as *hashtags*.

Baseado nos *links*, *hashtags* e no próprio texto de cada mensagem, uma arquitetura denominada *Dynamic Profile Mapper* (DPM) foi construída com o objetivo de montar dinamicamente três perfis distintos para cada usuário. São eles:

- **Perfil Texto** - É montado extraindo-se as entidades nomeadas relevantes somente do texto presente nas mensagens, não levando em consideração *links* ou *hashtags* inseridas.

- **Perfil *Link*** - É composto pelas entidades nomeadas extraídas dos *links* que são inseridos nas mensagens. Caso uma mensagem possua algum *link*, o mesmo é processado para que sejam extraídas entidades nomeadas de seu texto, porém caso haja algum *link* inserido nesse texto, o mesmo não é seguido para evitar caminhos longos ou até infinitos. Isso é feito para cada *link* de cada uma das mensagens inseridas.

- **Perfil *Trending Topic*** - É construído baseado nas *hashtags* inseridas nas mensagens, com o objetivo de descobrir o contexto da *hashtag*. Caso uma mensagem possua alguma *hashtag*, é feita uma busca pelos principais *posts* que possuam esta *hashtag* e o texto destes *posts* são compilados e processados para que as entidades nomeadas sejam extraídas. A mesma lógica aplicada para o perfil link que evita caminhos longos ou infinitos também é aplicada aqui, ou seja, caso haja outras *hashtags* inseridas no texto da *hashtag* original, as mesmas não são processadas.

Além da extração das entidades nomeadas que é realizada nos três tipos de perfis, os textos nos quais são realizadas estas extrações também são categorizados por tópicos. O DPM realiza a análise de cada texto tentando identificar algum tópico no qual o mesmo se enquadre e estes tópicos foram pré-definidos: artes e entretenimento, negócios, computadores e internet, cultura e política, jogos, saúde, temas policiais, religião, lazer, ciência e tecnologia, esportes, e previsão do tempo.

Com isso, esses tópicos também fazem parte dos três tipos de perfis construídos dos usuários.

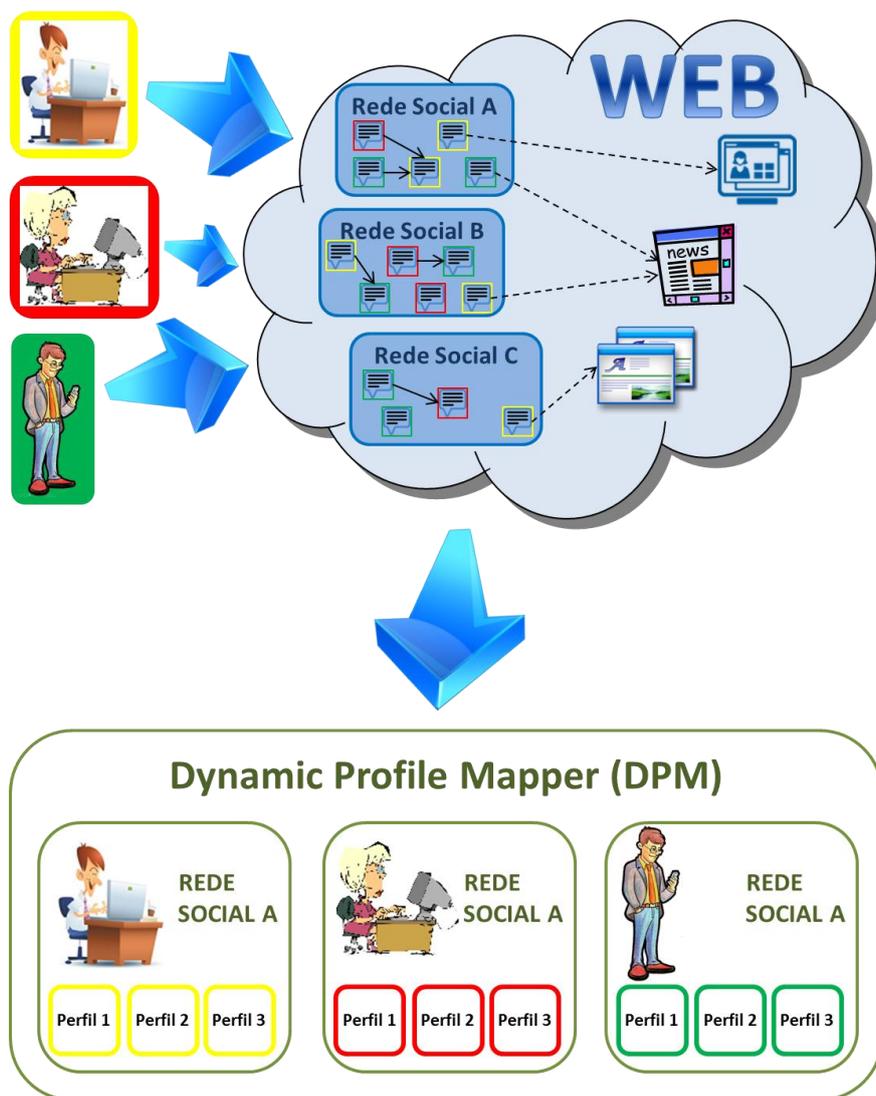


Figura 2 – Diagrama Conceitual

A Figura 2 ilustra o diagrama conceitual da proposta onde usuários realizam acesso à web pelos mais diversos meios e inserem mensagens em redes sociais, as quais são processadas para montar três perfis dinâmicos distintos para cada usuário. Estes perfis são construídos considerando as mensagens inseridas em cada rede social de forma separada.

3.2. Arquitetura lógica

Para poder processar cada mensagem inserida pelos usuários, juntamente com os elementos que a compõem, foi construída a arquitetura lógica apresentada na Figura 3.

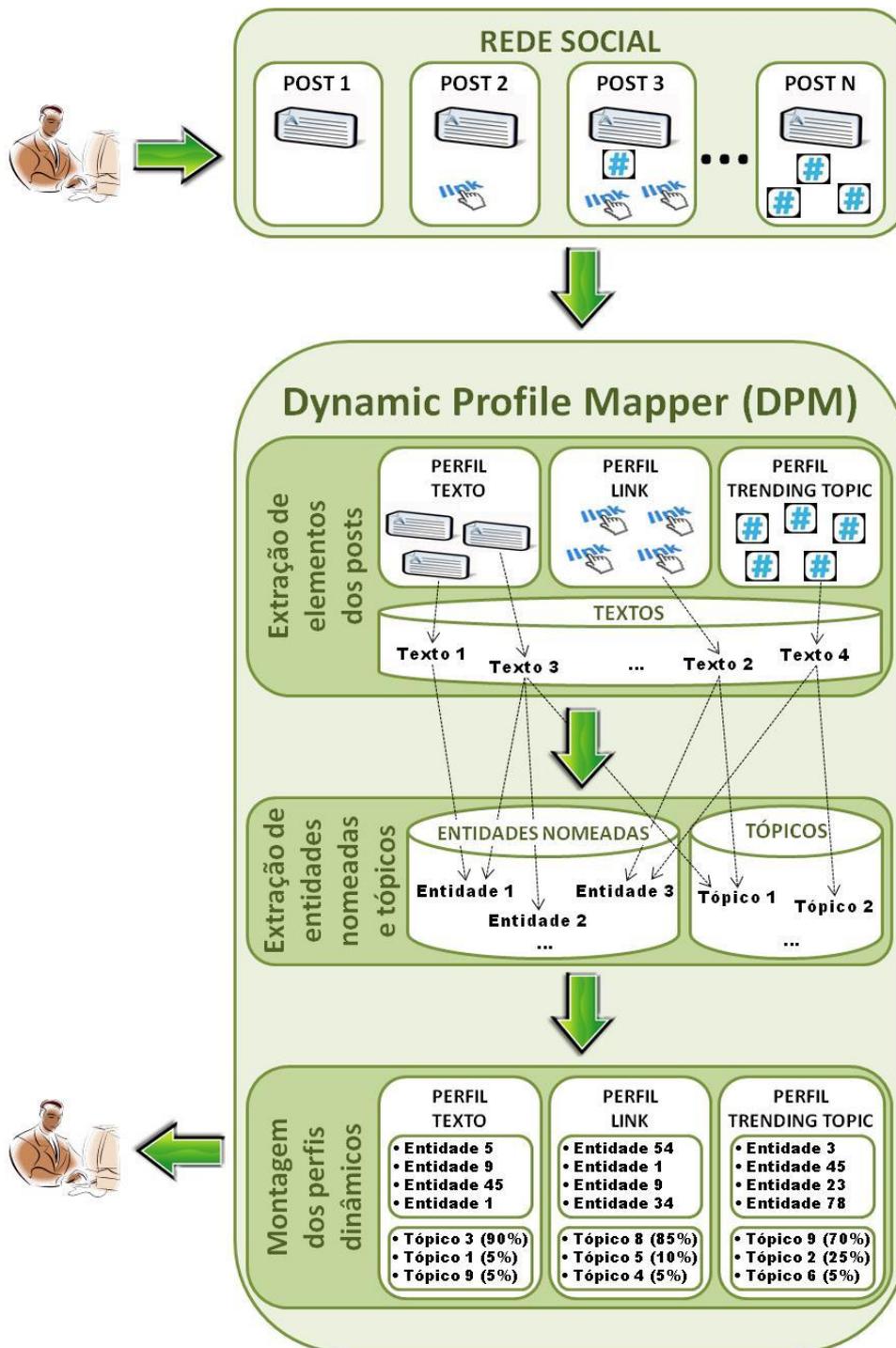


Figura 3 – Arquitetura lógica

A solução é dividida em três principais módulos: extração de elementos dos *posts*, extração de entidades nomeadas e tópicos, e montagem dos perfis dinâmicos. Todo o processo de construção dos perfis é realizado de forma individual, ou seja, cada usuário tem seus *posts* processados para montar seus respectivos perfis.

A seguir os três módulos serão detalhados.

3.2.1. Extração de elementos dos *posts*

Este módulo tem por objetivo recuperar todos os *posts* inseridos por um determinado usuário em uma rede social e processá-los para separar as informações pertinentes para cada tipo de perfil, ou seja, o texto dos *posts* é organizado para montar o perfil texto, o *link* dos *posts* serve para montar o perfil *link*, enquanto as *hashtags* montam o perfil *trending topic*.

O único parâmetro que este módulo recebe é o identificador do usuário na rede social em que se está trabalhando.

3.2.1.1. Montagem do perfil texto

Ao processar cada *post*, o DPM inicialmente busca o texto do mesmo desconsiderando todos os *links* e *hashtags* inseridos. Como este é o primeiro passo do processamento, a identificação do usuário, o identificador e a data de cada *post* também são armazenados.

É possível que não exista texto no *post* que está sendo processado, pois em redes sociais muitas vezes os usuários inserem somente links ou *hashtags* e nesses casos são armazenadas somente as informações genéricas deste *post*.

A Figura 4 ilustra um exemplo de um *post* considerado na montagem do perfil texto de um determinado usuário.



Figura 4 – Exemplo de montagem do perfil texto

3.2.1.2. Montagem do perfil *link*

Após a montagem do perfil texto, caso haja um ou mais *links* no *post*, estes *links* são armazenados e para cada um o DPM segue o *link* e extrai o corpo do texto deste *link*, mas caso este corpo possua outros *links*, esses são ignorados, ou seja, somente o texto do link presente no *post* é considerado. Entende-se por corpo do texto de uma página, o texto principal desta página, ou seja, somente são consideradas as informações relevantes do conteúdo. Informações como cabeçalho, rodapé, publicidade ou menu de navegação são desconsideradas por não agregarem conteúdo relevante a ser processado.

The image shows a screenshot of a news article on the website 'globoesporte.com'. The article is titled 'Zico apresenta projeto social em Votorantim nesta quinta-feira'. The main content area is annotated with a red box and an arrow pointing to it, labeled 'Conteúdo'. The top navigation bar and search area are annotated with a red box and an arrow, labeled 'Cabeçalho'. On the right side, there is a 'publicidade' section with a 'RE' logo and a list of sports news items, annotated with a red box and an arrow, labeled 'Publicidade'. At the bottom of the page, there is a navigation menu with categories like 'times série a', 'campeonatos', 'esportes', 'tv, impresso e rádio', and 'parcerias', annotated with a red box and an arrow, labeled 'Rodapé'. The footer contains copyright information for 2012 and links to 'globo.com', 'noticias', 'esportes', 'entretenimento', and 'videos'.

Figura 5 – Página contendo o conteúdo que será considerado no processamento

A Figura 5 ilustra um exemplo do conteúdo que é filtrado no processamento de uma página que é seguida através de um *link* inserido no *post*.

O texto do conteúdo é armazenado para posterior processamento onde serão extraídas as entidades nomeadas e também ocorrerá a identificação do tópico no qual este texto se enquadra.

A Figura 6 ilustra um exemplo de como é construído o perfil *link* através de um *post* que possui um *link* inserido.

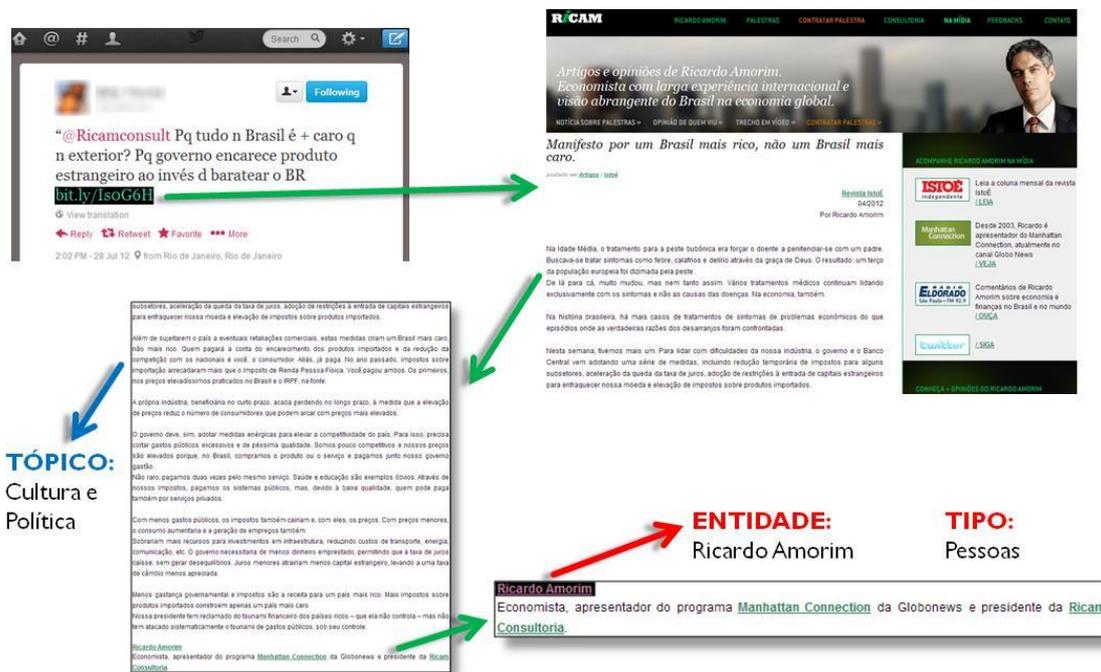


Figura 6 – Exemplo de montagem do perfil *link*

3.2.1.3. Montagem do perfil *trending topic*

O último processamento deste módulo ocorre caso haja alguma *hashtag* inserida no *post* que está sendo processado.

Seu funcionamento é similar ao perfil *link* no que tange à busca de conteúdo que será processado posteriormente, mas a grande diferença está como este conteúdo é construído.

Considerando que as *hashtags* servem prioritariamente como uma agregação de mensagens que se referem a um mesmo assunto, o DPM as utiliza para realizar uma busca por *posts* que também possuem esta mesma *hashtag* na rede social em que se está trabalhando e agrega todos esses textos em um só para que sejam extraídas as entidades nomeadas e ocorra a identificação do tópico onde o texto de enquadra no próximo módulo.

A Figura 7 ilustra um exemplo de busca por uma *hashtag* em uma rede social.



Figura 7 – Resultado da busca em uma rede social pela *hashtag* #redesocial

A Figura 8 ilustra um exemplo de como é construído o perfil *trending topic* através de um *post* que possui uma *hashtag* inserida.

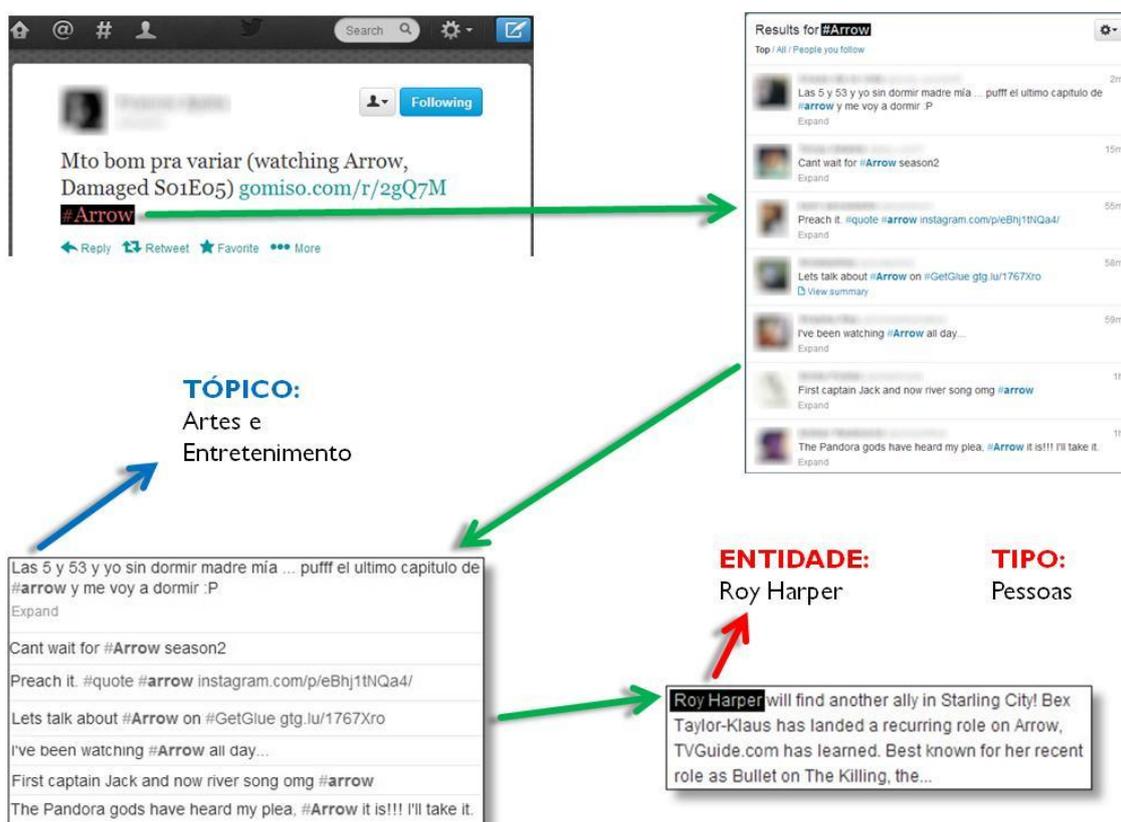


Figura 8 – Exemplo de montagem do perfil *trending topic*

3.2.2. Extração de entidades nomeadas e tópicos

Com todos os elementos organizados em seus respectivos perfis, o DPM inicia o processamento de todos os textos armazenados com o objetivo de extrair as entidades nomeadas de cada um, além de identificar o tópico no qual estes textos se enquadram.

Todas as entidades nomeadas extraídas são armazenadas em um repositório compartilhado por todos os tipos de perfis para evitar duplicidade de entidades. Apesar das entidades serem comuns a todos os tipos de perfis, este módulo tem a preocupação de guardar informações importantes de uma determinada entidade para cada texto onde ela se fez presente.

As informações obtidas de cada entidade são: (i) a relevância, que é representada por um valor numérico entre zero e um, sendo um a maior relevância possível, (ii) a quantidade de vezes que a entidade apareceu em cada texto, representada por um valor numérico inteiro positivo e (iii) o tipo da entidade que se enquadra nos seguintes tipos: organizações, pessoas, lugares, eventos e outros, sendo esta última categoria uma agregação de entidades que não se encaixam nas categorias anteriores.

Além da extração das entidades nomeadas de cada texto, é feita também uma análise desse texto com a finalidade de tentar descobrir algum tópico no qual ele se enquadre.

Caso algum texto não se enquadre em um desses tópicos ele ficará sem classificação, porém suas entidades nomeadas serão processadas normalmente, ou seja, a atividade de atribuição de um tópico a um texto independe da tarefa de extração de entidades nomeadas.

A mesma preocupação de guardar informações importantes de uma determinada entidade para cada texto onde ela se fez presente também é aplicada para os tópicos.

3.2.3. Montagem dos perfis dinâmicos

Para representar documentos ou consultas em um espaço vetorial, é necessário o cálculo de pesos para cada termo presente, sendo que este valor deve depender do número de ocorrências do termo nos documentos.

Após todas as entidades serem extraídas no módulo de extração de entidades nomeadas e tópicos, é preciso realizar um cálculo de pesos para definir em cada tipo de perfil quais foram as entidades nomeadas que melhor representaram cada usuário.

A montagem de cada perfil realizada pelo módulo de montagem dos perfis dinâmicos considerou como peso a frequência dos termos nos documentos processados sendo denotado por $TF_{t,d}$, sendo t o termo e d o documento.

Utiliza-se o TF para evidenciar os termos mais relevantes e que melhor representem um documento a partir do número de ocorrências de termos neste documento. Chegou-se a cogitar o uso de outra forma de cálculo de pesos muito utilizada, o TF-IDF que tenta diferenciar um termo dentro de uma coleção para demonstrar sua importância considerando uma lista de documentos que se está trabalhando. Esta medida aumenta se o termo ocorre em um documento e diminui se ocorre na lista de documentos com grande frequência, denotando assim sua originalidade. É bastante utilizado para desconsiderar, por exemplo, *stop words* que são palavras que podem ser consideradas irrelevantes para o conjunto de resultados que se está trabalhando. Alguns exemplos de *stop words* são: as, e, os, de, para, com, sem, foi. Como no DPM as entidades nomeadas são tratadas e já retornadas pela API de processamento de linguagem natural, não foi necessário preocupar-se com o tratamento de termos irrelevantes como *stop words*, por exemplo.

Em relação aos tópicos identificados de cada texto, é feito um cálculo percentual para consolidá-los e enquadrá-los em cada um dos três tipos de perfis construídos dos usuários.

Com os perfis montados pelo DPM, os usuários têm a possibilidade de verificar o que foi inferido em cada um de seus perfis e realizar consultas nessas informações.

3.3. Arquitetura física

A proposta de arquitetura lógica foi implementada na arquitetura física mostrada na Figura 9. Toda a parte de recuperação das informações de cada usuário para montar

seus perfis no DPM foi desenvolvida na linguagem de programação Java e todos os componentes utilizados nesta implementação são soluções baseadas em código aberto (*open-source*).

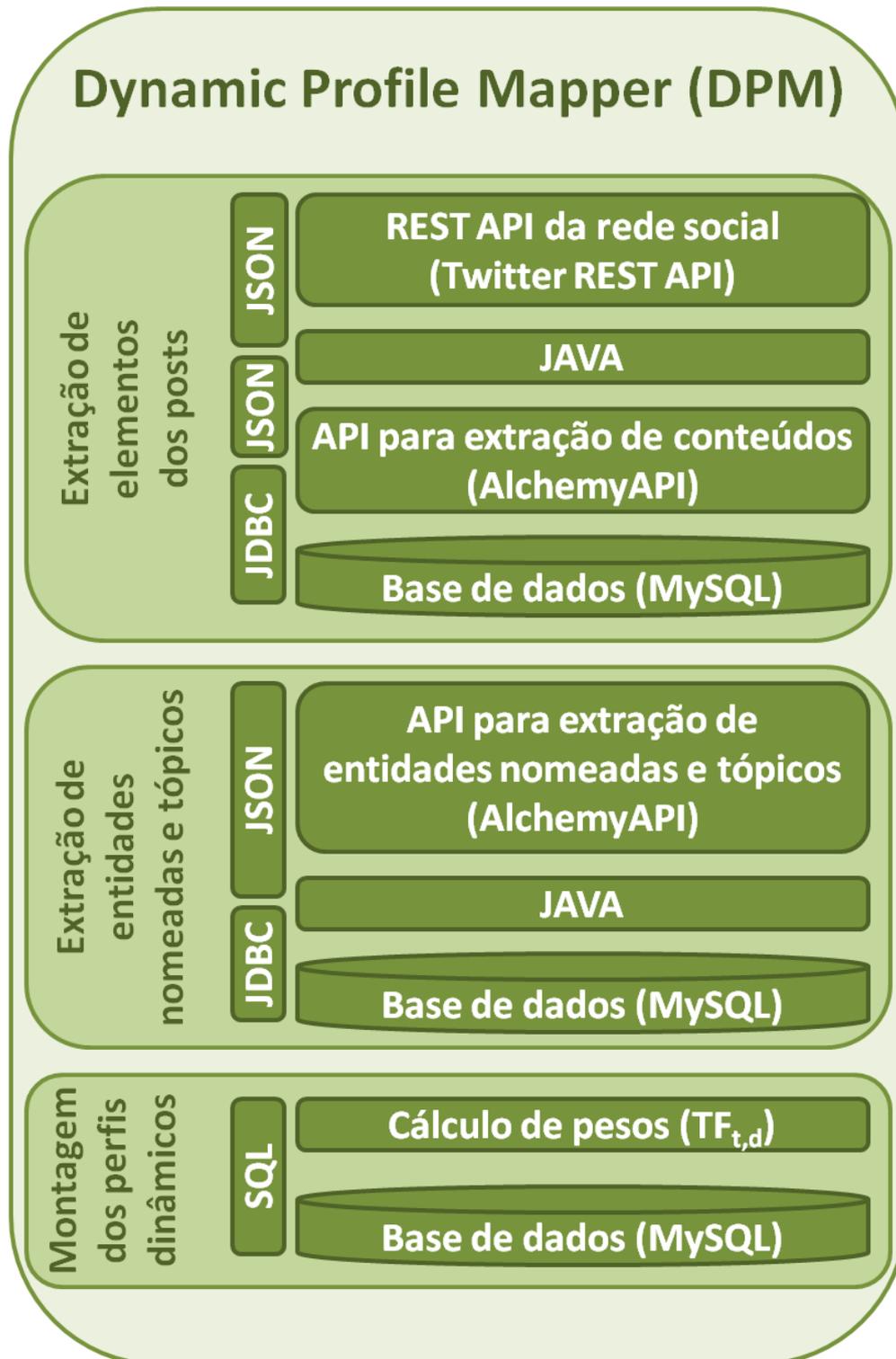


Figura 9 – Arquitetura física

3.3.1. Extração de elementos dos *posts*

Neste módulo, o DPM busca as informações de cada usuário, juntamente com os seus *posts* através da API (*application programming interface*) fornecida pela própria rede social objeto de análise através da tecnologia REST (*Representational State Transfer*).

REST é um termo definido por FIELDING (2000) para descrever um estilo arquitetural de sistemas de informação distribuídos:

"the name 'Representational State Transfer' is intended to evoke an image of how a well-designed web application behaves: a network of web pages (a virtual state-machine), where the user progresses through the application by selecting links (state transitions), resulting in the next page (representing the next state of the application) being transferred to the user and rendered for their use".
(FIELDING, 2000)

O modelo REST utiliza um conjunto de interfaces genéricas para promover interações sem estado (*stateless*) através da transferência de representações de recursos, em vez de operar diretamente sobre esses recursos (NUNES & DAVID, 2005).

As informações dos usuários são recuperadas através da API da rede social no formato JSON (*JavaScript Object Notation*) que é um formato leve e simples para troca de dados. A Figura 10 ilustra um trecho de dados representados em JSON.

```

1. [
2.   {
3.     "coordinates": null,
4.     "favorited": false,
5.     "truncated": false,
6.     "created_at": "Wed Aug 29 17:12:58 +0000 2012",
7.     "id_str": "240859602684612608",
8.     "entities": {
9.       "urls": [
10.        {
11.          "expanded_url": "https://dev.twitter.com/blog/twitter-certified-products",
12.          "url": "https://t.co/MjJ8xAnT",
13.          "indices": [
14.            52,
15.            73
16.          ],
17.          "display_url": "dev.twitter.com/blog/twitter-c\u2026"
18.        }
19.      ],
20.      "hashtags": [
21.
22.      ],
23.      "user_mentions": [
24.
25.      ]
26.    },
27.    "in_reply_to_user_id_str": null,
28.    "contributors": null,
29.    "text": "Introducing the Twitter Certified Products Program: https://t.co/MjJ8xAnT",
30.  }
31. ]

```

Figura 10 – Exemplo de dados representados em formato JSON

Os dados de cada usuário são recuperados de forma conjunta, trazendo informações de todos os perfis e, através de programação Java, os mesmos são separados e armazenados em uma base de dados MySQL⁷.

As informações dos perfis texto são as mais simples de serem tratadas, pois dizem respeito somente ao texto que foi inserido no *post* e por isso basta extrair essa informação da API e armazená-la diretamente na base de dados.

No caso do perfil *link* isso não é tão simples, pois a API da rede social retorna somente os *links* inseridos, sendo necessário seguir cada *link* e processá-los para extrair o seu conteúdo. Isso é feito através de uma API que realiza a extração de conteúdos da página, conforme ilustrado na Figura 4. A API utilizada para esta tarefa foi a AlchemyAPI⁸.

Esta API recebe como parâmetro uma URL (*Uniform Resource Locator*) e também retorna através de um objeto JSON o texto extraído da página. Em posse desse

⁷ Site: <http://www.mysql.com>

⁸ Site: <http://www.alchemyapi.com>

texto, o DPM armazena essas informações em base de dados. A utilização desta API em detrimento de outras como a Boilerpipe⁹ que também é utilizada em várias pesquisas para este fim, deveu-se ao fato da primeira ter se mostrado mais rápida no retorno dos resultados. Em relação à precisão dos resultados obtidos, as duas APIs mostraram-se equivalentes.

Em sua pesquisa, KOVAČIČ (2012) realiza uma série de comparações com algumas soluções para extração de conteúdos em páginas *web* e a AlchemyAPI apresentou bons resultados. Outro ponto que desfavoreceu a escolha da API Boilerpipe foi o fato da mesma não sofrer atualização desde junho de 2011. Isso é um ponto negativo, pois como a forma de codificação das páginas *web* estão em constante transformação, é preciso sempre rever as heurísticas utilizadas para extrair esses conteúdos. Essa situação não ocorre na AlchemyAPI, pois além de possuir cotas de utilização gratuitas (incluindo cotas extras para pesquisadores), também possui pacotes comerciais e por isso tem uma maior necessidade de manter-se atualizada.

Para o perfil *trending topic* é feito um tratamento distinto dos demais, pois para cada *hashtag* encontrada é feita uma busca por *posts* que também possuem esta mesma *hashtag* e todos esses textos são agregados para posterior extração das entidades nomeadas. Essa busca de *posts* com a mesma *hashtag* é feita também pela API da rede social, porém em um módulo diferente do que recupera as informações de usuários. Esta API também retorna os resultados em formato JSON e através de programação Java, todos esses textos são agregados e armazenados em base de dados.

3.3.2. Extração de entidades nomeadas e tópicos

Com todos os textos organizados e armazenados em base de dados para cada tipo

⁹ Site: <https://code.google.com/p/boilerpipe>

distinto de perfil, a fase de extração de entidades nomeadas e tópicos tem seu início. Para isso é necessária a realização do processamento de linguagem natural para que as entidades sejam reconhecidas e extraídas, além da identificação dos tópicos dos textos.

Para reconhecer entidades nomeadas, existe uma grande variedade de APIs disponíveis na *web*. Entre as mais utilizadas podemos citar o OpenCalais¹⁰, mas apesar de ser uma boa ferramenta, até a data da realização desta pesquisa, não possuía suporte para a língua portuguesa. Isso é muito importante, pois apesar de o DPM ser flexível o suficiente para suportar qualquer idioma, a língua portuguesa foi fundamental para a realização do estudo de caso já que é a língua nativa de todos os participantes do estudo de caso, como veremos no Capítulo 5.

Em função disso, três opções foram consideradas: F-EXT¹¹, Ltasks¹² e AlchemyAPI. A escolhida foi novamente a AlchemyAPI que, no caso de extração de entidades nomeadas, sua plataforma disponibiliza um conjunto de funcionalidades de processamento de linguagem natural e utiliza LOD¹³ (*Linked Open Data*) como fonte de conhecimento (LIMA *et al.*, 2012).

Em seu trabalho LIMA *et al.* (2012) fazem uma comparação entre APIs que realizam reconhecimento de entidades nomeadas na língua portuguesa e a AlchemyAPI foi apontada com a melhor acurácia e a menor taxa de erro entre as participantes do estudo. A não escolha da F-EXT e da Ltasks se deu por razões similares às citadas na seção anterior, ou seja, são duas APIs que não sofrem atualizações constantes, sendo que a primeira teve sua documentação atualizada pela última vez em 25 de março de 2010 e a segunda no dia 20 de setembro de 2011.

¹⁰ Site: <http://www.opencalais.com>

¹¹ Site: <http://www.learn.inf.puc-rio.br/fextws/index.jsf>

¹² Site: <http://ltasks.com>

¹³ Site: <http://linkeddata.org>

Outra forte razão de escolha da AlchemyAPI se deu ao fato da grande quantidade de idiomas que são disponibilizados por ela, apesar de somente as línguas portuguesa e inglesa terem sido consideradas na pesquisa, sendo que esta última se deve ao fato da grande quantidade de referências existentes neste idioma nos textos processados.

Importante ainda ressaltar que a AlchemyAPI também foi utilizada para identificação dos tópicos de cada texto.

Para cada texto processado pela API, todas as entidades nomeadas encontradas eram armazenadas na mesma base de dados MySQL utilizada no módulo de extração de elementos dos *posts*.

Cada entidade nomeada retornada possuía também um valor numérico representando sua relevância que era calculado pela API baseado na importância que a mesma tinha dentro do texto processado. Além disso, a API também tentava identificar um tópico no qual o texto se enquadrava e, junto deste tópico, também era retornado o seu valor de relevância. Essas informações também foram armazenadas no banco de dados e serviram de base para que o próximo módulo pudesse ter insumos para apontar as entidades mais representativas para cada tipo de perfil dos usuários, além dos tópicos que os usuários mais se interessavam.

3.3.3. Montagem dos perfis dinâmicos

Depois de todas as entidades nomeadas estarem armazenadas em base de dados, é chegado o momento de ranqueá-las calculando o peso de cada uma para cada tipo de perfil.

Isso é feito através de consultas na base de dados utilizando a linguagem SQL (*Structured Query Language*) que traz as entidades nomeadas mais representativas para cada tipo de perfil.

A ordenação destas entidades se dá através do número de ocorrências das mesmas nos textos dos *posts* inseridos pelos usuários. É importante lembrar que o texto no caso do perfil texto é o próprio texto da mensagem, excluindo-se os *links* e *hashtags* que porventura existam, no caso do perfil *link* é o conteúdo extraído da página o qual o mesmo aponta e no caso do perfil *trending topic* é a agregação dos textos que também possuem a mesma *hashtag* inserida. Quanto maior o número de ocorrências que uma entidade possua, melhor será o seu ranqueamento e mais representativa ela será em um determinado perfil para aquele usuário.

Caso haja um empate no número de ocorrências das entidades, o desempate é feito através de um cálculo da média das relevâncias que cada entidade obteve em cada um dos textos aos quais ela se fez presente. Exemplificando para facilitar o entendimento: as entidades A e B aparecerem duas vezes em textos distintos para um determinado perfil de um usuário, sendo que a entidade A obteve 90% relevância no primeiro texto e 70% no segundo. Já a entidade B obteve 80% de relevância no primeiro texto e 86% no segundo. Logo, o ranqueamento da entidade B será melhor que o da entidade A, pois B obteve média de 83% de relevância, enquanto A obteve 80%.

No caso dos tópicos que os usuários mais se interessavam, a linguagem SQL também foi utilizada para recuperá-los da base de dados e o cálculo foi feito baseado no valor de relevância de cada tópico inferido para cada texto e somente os tópicos que possuíssem uma relevância acima de 50% foram considerados em cada tipo de perfil dos usuários. Isso pode ser explicado pelo fato de se tentar aumentar a representatividade dos tópicos aos quais os usuários mais se interessavam em cada tipo de perfil.

Capítulo 4 – Estudo de Caso

Neste capítulo serão descritos o método da pesquisa, a dinâmica das atividades realizadas no estudo de caso e a caracterização dos participantes.

4.1. Metodologia

O método mais apropriado para ser utilizado na pesquisa é o estudo de caso, pois é aplicável em casos onde o pesquisador tem pouco controle sobre os eventos e o foco se encontra em fenômenos contemporâneos inseridos em algum contexto da vida real (YIN, 2005). Segundo Yin, estudo de caso se baseia nas características do fenômeno em estudo e com base num conjunto de características associadas ao processo de obtenção de dados e às estratégias de análise dos mesmos. É uma abordagem metodológica de investigação adequada quando se procura explorar, descrever ou compreender acontecimentos e contextos complexos, onde diversos fatores estão simultaneamente envolvidos.

Algumas características que um estudo de caso deve possuir, segundo (BENBASAT *et al.*, 1987):

- Fenômeno observado no seu ambiente natural;
- Os dados podem ser obtidos utilizando-se diversos meios, tais quais: observações diretas e indiretas, entrevistas, questionários, registros de áudio e

vídeo, diários, cartas, etc.;

- Uma ou mais entidades (pessoa, grupo, organização) são analisadas;
- A complexidade da unidade é estudada de forma aprofundada;
- Pesquisa dirigida aos estágios de exploração, classificação e desenvolvimento de hipóteses do processo de construção do conhecimento;
- Não são utilizadas formas experimentais de controle ou manipulação;
- O investigador não precisa especificar antecipadamente o conjunto de variáveis dependentes e independentes;
- Podem ser feitas mudanças na seleção do caso ou dos métodos de obtenção de dados à medida que o investigador desenvolve novas hipóteses;

O estudo de caso neste trabalho foi realizado através de uma pesquisa qualitativa realizada com usuários que utilizam redes sociais.

O caráter da pesquisa qualitativa usada nesse trabalho é essencialmente exploratório. Esse perfil de pesquisa foi escolhido devido à flexibilidade, criatividade e informalidade que ele permite ao pesquisador que busca um maior conhecimento sobre o tema ou problema de pesquisa (AAKER *et al.*, 1995).

De acordo com KINNEAR & TAYLOR (1996) este tipo de estudo é apropriado quando os objetivos de pesquisa incluem situações onde temos a necessidade de:

- Desenvolver uma formulação mais precisa de um problema ainda vagamente identificado;
- Familiarizar e elevar o conhecimento e a compreensão de um problema;
- Ganhar perspectiva no que diz respeito à gama de variáveis que existem em um determinado problema de pesquisa;
- Estabelecer prioridades em relação à significância das várias variáveis do problema;

- Identificar e formular cursos de ação para o estudo realizado.

Na tabela abaixo é possível verificar as principais diferenças entre pesquisas exploratórias e conclusivas:

Tabela 2 - Diferenças entre pesquisa exploratória e conclusiva (PARASURAMAN, 1991)

Componentes do Projeto de Pesquisa	Pesquisa exploratória	Pesquisa conclusiva
Propósito da Pesquisa	Geral: para gerar esclarecimento sobre uma situação.	Específico: para gerar esclarecimento e ajudar a selecionar um curso de ação.
Necessidade de dados	Vaga	Clara
Origem dos dados	Parcialmente definida	Bem definida
Forma de coleta de dados	Aberta-fechada, Flexível	Bem definida
Amostra	Relativamente pequena; selecionada subjetivamente para maximizar a geração de informações utilizáveis.	Relativamente ampla, selecionada objetivamente para permitir a generalização dos achados.
Coleta de dados	Flexível, sem procedimento estruturado.	Rígida com procedimento bem estruturado.
Análise de dados	Informal, tipicamente não quantitativa.	Formal, tipicamente quantitativa.

Enquanto estudos quantitativos geralmente procuram seguir com rigor um plano previamente estabelecido (baseado em hipóteses claramente indicadas e variáveis que são objeto de definição operacional), a pesquisa qualitativa costuma ser direcionada, ao longo de seu desenvolvimento. Além disso, não busca enumerar ou medir eventos e, geralmente, não emprega instrumental estatístico para análise dos dados. Seu foco de interesse é amplo e parte de uma perspectiva diferenciada da adotada pelos métodos quantitativos. Dela faz parte a obtenção de dados descritivos mediante contato direto e interativo do pesquisador com a situação objeto de estudo. Nas pesquisas qualitativas, é frequente que o pesquisador procure entender os fenômenos, segundo a perspectiva dos participantes da situação estudada e, a partir, daí situe sua interpretação dos fenômenos estudados (NEVES, 1996).

Segundo GODOY (1995), existe um conjunto de características essenciais capazes de identificar uma pesquisa qualitativa: (i) utiliza um ambiente natural como

fonte direta de dados e o pesquisador como instrumento fundamental; (ii) possui caráter descritivo; (iii) possui um enfoque indutivo.

Para conduzir a pesquisa qualitativa, foi utilizado o método MEDS (Método de Extração do Discurso Subjacente) (NICOLACI-DA-COSTA *et al.*, 2004). O MEDS é um método exploratório desenvolvido para a pesquisa em psicologia clínica e tem em comum com os demais métodos qualitativos a abertura, a investigação aprofundada em contexto, as amostras pequenas e a flexibilidade de procedimentos e técnicas. É um método fortemente interdisciplinar. Seu desenvolvimento sofreu uma grande influência da linguística, das teorias da construção social da subjetividade, bem como de alguns pressupostos e técnicas de investigação clínica oriundos da psicanálise.

A principal diferença entre o MEDS e grande parte dos demais métodos qualitativos é o registro explícito de que trabalha com material discursivo, aliado a uma concepção de que o discurso desempenha um importante papel na construção social da esfera do psicológico (NICOLACI-DA-COSTA *et al.*, 2004).

A coleta de dados é tradicionalmente realizada por meio de entrevistas presenciais que devem se aproximar o máximo possível de conversas informais cotidianas e esta é uma das grandes preocupações do MEDS. Para garantir essa informalidade, o MEDS adota diversos procedimentos como: (i) ter como base um roteiro de itens a partir do qual as perguntas sejam formuladas durante a própria entrevista de modo que não soem artificiais; (ii) respeitar o fluxo do raciocínio do entrevistado por meio da aplicação flexível desse roteiro (o que significa dizer que a ordem das perguntas pode ser alterada); e (iii) conduzir as entrevistas em contextos informais nos quais os entrevistados se sintam à vontade. Todos esses procedimentos também se aplicam às entrevistas on-line (NICOLACI-DA-COSTA, 2009). Em função dos participantes da pesquisa se tratar de indivíduos que passam grande parte de seus dias na *web* e por isso

possuem grande fluência no uso de tecnologias de comunicação, caso dos mensageiros instantâneos, optou-se pela realização de entrevistas on-line.

Tanto nas entrevistas presenciais quanto nas entrevistas on-line, o MEDS tende a buscar a homogeneidade e este é o caso desta pesquisa conforme será descrito na seção 4.3.2.

Existe consenso de que as pesquisas qualitativas são bastante trabalhosas, demandam muito tempo e em função disso usam poucos participantes (RUBIN & RUBIN, 2005; TAYLOR, 2001; WEISS, 1994). Além disso, o número de participantes raramente é estipulado inicialmente, pois o principal critério usado para determinar se as entrevistas realizadas são suficientes para a investigação de um determinado assunto é o fenômeno conhecido como “ponto de saturação” (SEIDMAN, 1998; TURATO, 2003).

O ponto de saturação ocorre quando, após um certo número de entrevistas, há uma rarefação de informações novas e o entrevistador começa a ouvir (ou ler, no caso das entrevistas on-line) relatos muito semelhantes àqueles que já ouviu (ou leu) feitos por novos entrevistados. Isso ocorre tanto presencialmente quanto on-line (NICOLACI-DA-COSTA, 2009).

No que diz respeito à construção do roteiro, NICOLACI-DA-COSTA (2009) cita que o MEDS trata da mesma forma as entrevistas presenciais e on-line. Nos dois casos é detalhista, pois adota o ponto de vista de que um bom roteiro é indispensável para uma boa pesquisa e estipula diretrizes gerais para sua construção. São elas:

- os roteiros devem ser estruturados em sua concepção, porém aplicados com flexibilidade e devem se inspirar em conversas naturais;
- para que as perguntas sejam automaticamente adaptadas ao contexto das entrevistas, preservando todas as suas características linguísticas e extralinguísticas naturais, os roteiros devem constar apenas de itens que serão

expandidos sob a forma de questões durante as próprias entrevistas e no caso das entrevistas on-line, a formulação das questões deve respeitar todas as formas de escrever características dos ambientes interativos on-line (CRYSTAL, 2006);

- os itens devem gerar perguntas abertas que comportem qualquer tipo de resposta (como, por exemplo, “O que você acha de X?”, “O que Y gera em você?”);
- perguntas de esclarecimento e/ ou aprofundamento (como “por quê?”, “como?”, “dá para explicar melhor?”, etc.) também devem ser previstas, mas podem ser introduzidas a qualquer momento em que sejam consideradas necessárias;
- para preservar a naturalidade de uma conversa informal, que nunca é composta somente de perguntas abertas, alguns itens deverão gerar perguntas fechadas cujas respostas são sim ou não, seguidas de perguntas de esclarecimento e/ ou aprofundamento;
- itens que geram perguntas mais abstratas, as quais solicitam opiniões, reflexões, posturas, sentimentos, avaliações, etc. do entrevistado a respeito de determinados tópicos, devem poder ser confrontados com itens que geram informações objetivas a respeito dos mesmos tópicos. Isso porque o contraste entre informações concretas e afirmações abstratas tem o potencial de revelar inconsistências que podem ser importantes vias de acesso ao discurso subjacente.

Após as entrevistas serem realizadas, é necessário trabalhar e analisar o material coletado. Tanto no caso dos depoimentos coletados presencialmente quanto no caso

daqueles coletados on-line, o MEDS tem como objetivo sua interpretação e não a verificação de hipóteses (NICOLACI-DA-COSTA, 2009).

A análise do MEDS é realizada em duas etapas complementares. A primeira, chamada de inter-participantes, consiste na análise das respostas dadas pelo grupo como um todo, enquanto a segunda, denominada intra-participantes, se dedica à análise detalhada de cada uma das entrevistas individuais (NICOLACI-DA-COSTA, 2009). Nesta pesquisa as duas análises foram realizadas e seus resultados são apresentados de forma conjunta.

4.2. Preparação do Ambiente

Inicialmente foi necessária uma pesquisa intensa na rede social Twitter por usuários que preenchiam o requisito mínimo necessário (detalhado na seção 4.3.2) para ter seus dados coletados e conseqüentemente estarem aptos a participar da pesquisa.

O Twitter foi criado no ano de 2006 e conta com mais de 500 milhões de usuários, sendo 170 milhões ativos (TECHCRUNCH, 2012). A escolha desta rede social deve-se ao grande desafio de buscar informações dos usuários com *posts* que contém um número muito pequeno de caracteres, no máximo 140.

Em um segundo momento, houve uma coleta de dados dos usuários elegíveis a participar da pesquisa e todos estes dados foram armazenados em base de dados seguindo toda a mecânica explicada no capítulo 3. Ao todo foram coletados dados de 15 usuários, porém não houve a necessidade de que todos fossem entrevistados devido ao atingimento do ponto de saturação que será explicado na seção 5.3 do Capítulo 5.

Com esses dados organizados e armazenados em base de dados, foi implementado um protótipo para que fosse possível realizar o estudo de caso e o mesmo foi

disponibilizado em um servidor do grupo de pesquisa Semantics & Learning (SaL)¹⁴ da UNIRIO. Alguns testes foram realizados para certificar que o protótipo estava funcional e um endereço público para acesso ao protótipo foi disponibilizado para os usuários participantes da pesquisa de forma que cada usuário só tivesse acesso aos dados inferidos do seu próprio perfil.

Estando o protótipo configurado, testado e disponível para acesso público, todos os participantes elegíveis foram convidados a participar da pesquisa através de e-mails e mensagens enviadas via Twitter e Facebook. Estas mensagens continham uma breve explicação do que se tratava a pesquisa, informava também que um protótipo havia sido construído para exibir os dados inferidos dos seus perfis, disponibilizava o *link* (somente do usuário que estava recebendo a mensagem) do servidor para acesso aos dados do seu perfil, os convidava para participar da pesquisa e, por último, caso aceitassem, era explicado que seria realizada uma entrevista on-line que durava cerca de 45 minutos, de forma a se obter os dados necessários para posterior análise do pesquisador.

Caso os usuários aceitassem participar da pesquisa, era disponibilizado o tempo que fosse necessário para que os mesmos pudessem explorar livremente o protótipo e depois que os usuários sinalizavam que estavam aptos a realizar a entrevista, a mesma era agendada de acordo com a disponibilidade do próprio usuário e do pesquisador.

4.3. Realização do Estudo de Caso

Para a realização do estudo de caso foi necessária a construção de um protótipo considerando as informações dos usuários obtidas no Twitter para que os mesmos pudessem observar e analisar os dados inferidos de seus perfis.

¹⁴ Site: <http://sal.uniriotec.br>

Os usuários participantes do estudo de caso foram selecionados seguindo o critério de perfil ativo explicado na seção 4.3.2.

4.3.1. Protótipo

Para avaliação do modelo proposto, foi construído um protótipo que organizou em três perfis distintos todas as informações obtidas do Twitter. Estes dados foram obtidos no ano de 2012 (de 01/01/2012 até 31/12/2012). Juntos, todos os usuários participantes da pesquisa geraram 5.097 *posts*, os quais muitos deles continham diversos *links* e *hashtags*.

Para coletar as mensagens inseridas pelos usuários e todas as informações contidas nas mesmas, foi utilizada a API do Twitter¹⁵. Através da linguagem Java, todas as informações obtidas no formato JSON foram organizadas e armazenadas na base de dados MySQL.

O protótipo consiste em uma série de páginas construídas através de *servlets* Java que proporcionam a possibilidade de adicionar conteúdo dinâmico em um servidor *web* usando a plataforma Java, interpretadas através do Apache Tomcat¹⁶ e traduzidas para o formato HTML (*Hyper Text Markup Language*) onde os usuários participantes da pesquisa puderam consultar todas as informações de seus três tipos de perfis, sendo possível, inclusive, filtrar as informações por intervalos de tempo pré-definidos (quadrimestrais), além de filtrar pelas categorias disponíveis das entidades nomeadas.

Inicialmente o usuário acessa uma tela que explica sucintamente a forma de construção de cada tipo de perfil e instrui este usuário a selecionar o tipo de perfil, juntamente com os filtros adicionais que se encontram no topo da página, conforme mostra a Figura 11.

¹⁵ Site: <https://dev.twitter.com>

¹⁶ Site: <http://tomcat.apache.org>

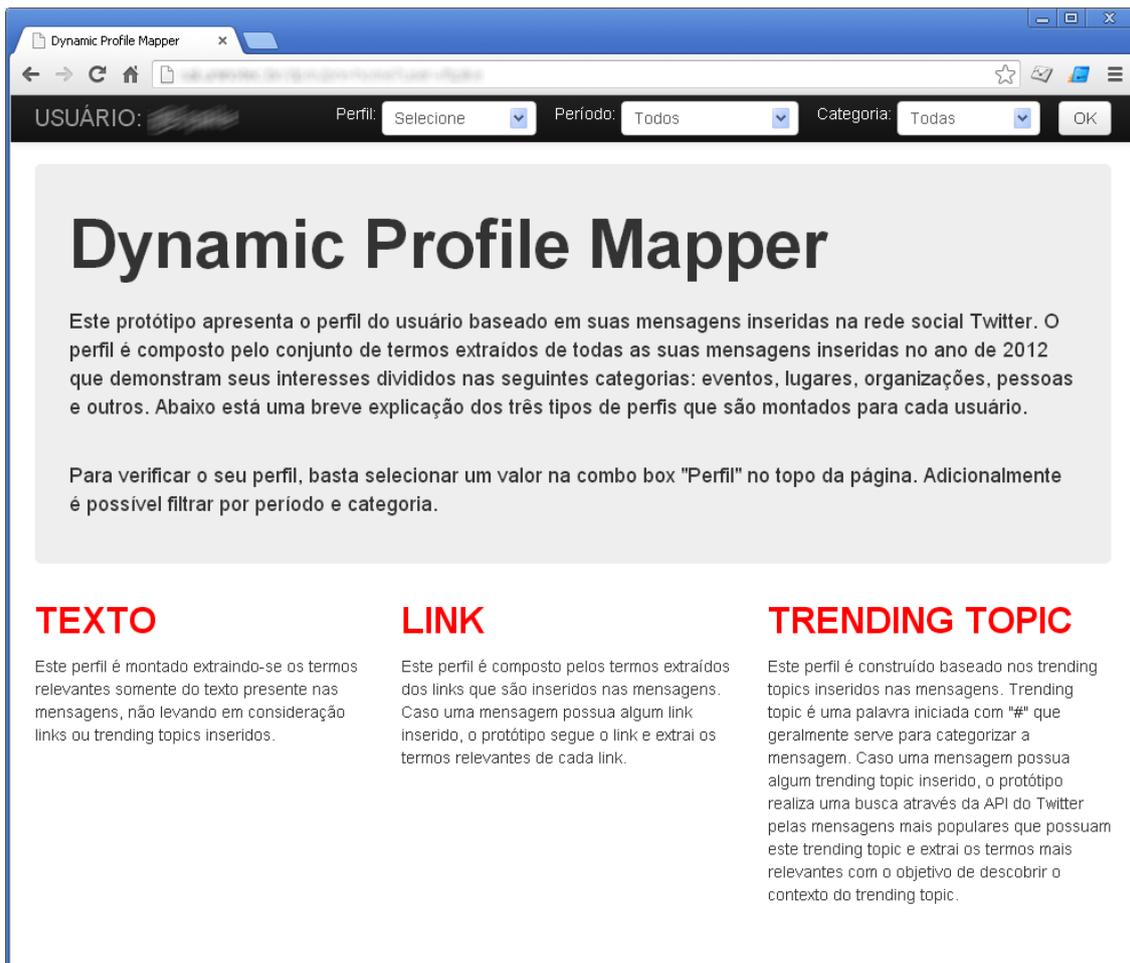


Figura 11 – Tela inicial contendo breve explicação sobre a montagem dos perfis

À medida que o usuário modifica os filtros adicionais juntamente com o perfil, o protótipo exibe um gráfico com os tópicos inferidos dos *posts* que se enquadram nos filtros de busca, além das entidades nomeadas, conforme ilustra a Figura 12. No protótipo os tópicos foram chamados de “Assuntos” e as entidades nomeadas chamadas de “Interesses” para facilitar a compreensão dos usuários.

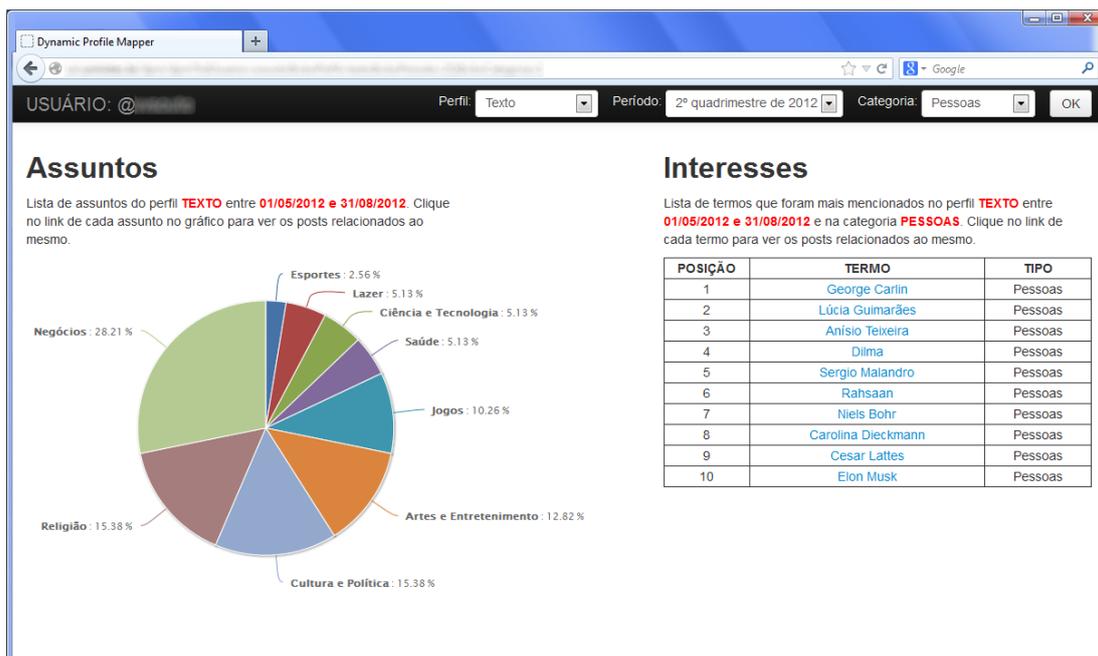


Figura 12 – Tela de consulta de informações mostrando as entidades nomeadas do tipo “pessoas”, no 2º quadrimestre de 2012 e no perfil TEXTO

Para cada entidade nomeada (interesse) apresentada no perfil do usuário e para cada tópico (assunto) é possível realizar o rastreamento dos *posts* que deram origem aos mesmos como mostra a Figura 13. Para isso, basta selecionar um dos *links* da tabela de interesses ou selecionar uma das categorias do gráfico de assuntos.



Figura 13 – Tela que exibe os *posts* relacionados à entidade nomeada “JEFF BEZOS”, no 3º quadrimestre de 2012 e no perfil TEXTO

Com a tela dos *posts* relacionados aberta, é possível ainda selecionar qualquer um dos *posts* listados na Figura 10 para que o mesmo seja exibido na própria interface do Twitter, conforme ilustrado na Figura 14.



Figura 14 – Post sendo exibido no Twitter

Para explicitar como a entidade nomeada foi extraída de um *post* que continha um *link* ou uma *hashtag*, é possível clicar em um ícone presente ao lado de qualquer *post* presente na lista de *posts* da Figura 13 e a tela presente na Figura 15 é exibida com a entidade nomeada sendo grifada todas as vezes que aparece no texto. Este ícone só é exibido para os casos dos perfis *link* e *trending topic*, pois no caso do perfil texto, o próprio texto que contém a entidade nomeada já está sendo exibido.

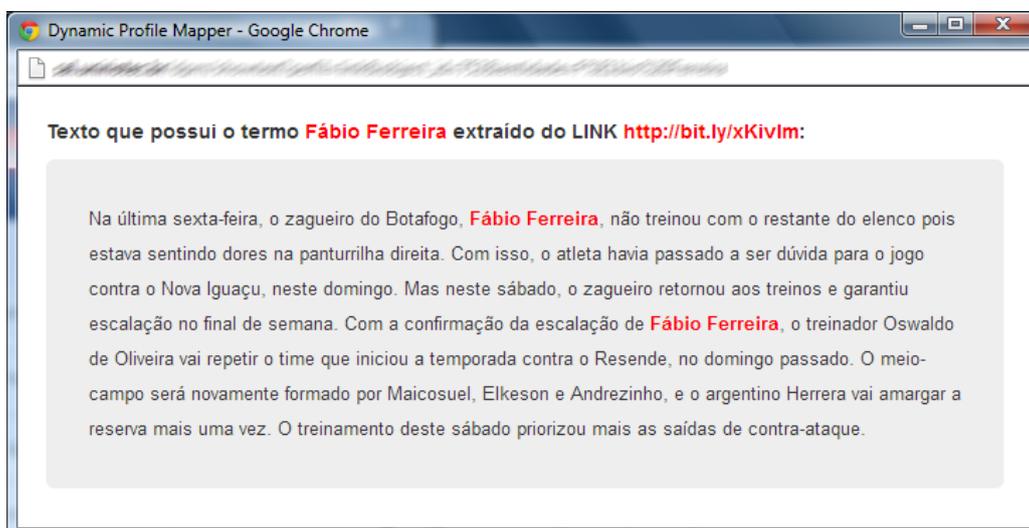


Figura 15 – Exibição do texto extraído de um link

4.3.2. Perfil dos Participantes

Dentre os 15 participantes que tiveram seus dados coletados, o estudo de caso considerou 10 que possuíam um perfil de publicação ativo na rede social Twitter no ano de 2012.

Entende-se por perfil ativo, perfis que tenham em média uma mensagem inserida a cada dois dias úteis do ano e, como em um ano o número de dias úteis gira em torno de 250, somente usuários com no mínimo 125 mensagens estariam elegíveis a participar da pesquisa. Isso se justifica pela importância de se coletar informações suficientes de cada usuário para montar o seu perfil dinâmico e com poucas mensagens acredita-se que esse trabalho ficaria inviável.

Todos os participantes eram do sexo masculino e possuíam idade entre 29 e 43 anos. A Figura 16 exibe mais alguns dados obtidos dos participantes durante a execução das entrevistas.



Figura 16 – Informações sobre os participantes das entrevistas

Capítulo 5 – Avaliação

Para avaliar a pesquisa, um questionário com 13 perguntas foi elaborado para que os usuários pudessem respondê-lo e passarem suas impressões sobre o que foi inferido em seus perfis construídos.

Para responder o questionário, foram realizadas entrevistas on-line através de meios eletrônicos como Google Talk ¹⁷, Facebook Chat ¹⁸ e Microsoft Office Communicator¹⁹ e as entrevistas duraram de 40 a 60 minutos.

O MEDS, assim como outros métodos qualitativos, tem um objetivo genérico: o de ouvir detalhadamente (a profundidade é um alvo comum) aquilo que, em contextos naturais e da forma mais livre possível, os entrevistados têm a dizer (NICOLACI-DA-COSTA, 2007).

Antes de realizar as entrevistas que serão consideradas e avaliadas na pesquisa, o MEDS cita que sejam realizadas entrevistas-piloto onde é possível obter vários ganhos, tais quais: (i) treinar o entrevistador inexperiente, (ii) testar a adequação do roteiro ao objetivo da pesquisa, pois muitas vezes julgamos que o roteiro está adequado e este teste mostra que não está e (iii) testar a sua fluidez, já que um bom roteiro é aquele que flui de um item a outro numa lógica análoga à de uma conversa informal (NICOLACI-DA-COSTA *et al.*, 2004).

¹⁷ Site: <http://www.google.com/talk/intl/pt-BR/index.html>

¹⁸ Site: <https://www.facebook.com/sitetour/chat.php>

¹⁹ Site: <http://www.microsoft.com/pt-br/download/details.aspx?id=8468>

O número de entrevistas-piloto necessárias depende de quanto o entrevistador está treinado e também da qualidade do roteiro elaborado. Quando as entrevistas fluem bem, é chegada a hora de encerrar esta etapa e começar as entrevistas que serão consideradas na pesquisa.

Foi realizada uma entrevista-piloto com um dos usuários elegíveis a participar da pesquisa e após a entrevista, ocorreu uma pequena mudança no roteiro que foi a inversão de ordem de duas perguntas, ou seja, uma mudança pouco significativa. Apesar do MEDS sugerir que entrevistas-piloto não façam parte do conjunto de entrevistas a ser analisado, decidiu-se que todos os dados obtidos nesta entrevista inicial também fossem considerados na análise, visto que a alteração no roteiro das entrevistas não causou perda de informação relevante a ser analisada.

Após todas as entrevistas terem sido executadas seguindo o roteiro final obtido através da entrevista-piloto, teve início a fase de análise dos dados chamadas de inter-participantes e intra-participantes. As duas análises serão detalhadas a seguir.

5.1. Análise inter-participantes e intra-participantes

Com os dados das entrevistas, seguindo o fluxo de trabalho citado pelo MEDS, é chegada a hora de realizar as análises inter-participantes e intra-participantes. A primeira tem o papel de fornecer ao pesquisador uma visão geral dos resultados obtidos e para isso é muito importante que as respostas dadas pelos participantes para cada pergunta sejam interpretadas de forma conjunta, além de serem sistematicamente comparadas em busca de recorrências (NICOLACI-DA-COSTA, 2007). Esta visão geral é aprofundada na análise intra-participantes que tem por objetivo analisar detalhadamente as respostas fornecidas por cada um dos participantes.

As análises inter e intra-participantes podem se suceder, em um processo iterativo, até que o material coletado seja apreendido em toda a sua complexidade (NICOLACI-DA-COSTA, 2007).

As duas análises serão apresentadas em conjunto, sendo divididas por perguntas ou grupo de perguntas que constam nas entrevistas. Será apresentada uma visão geral das respostas coletadas dos participantes, assim como alguns trechos das entrevistas.

5.1.1. Como aprimorar a abordagem e a consulta das informações

No início do questionário existiam perguntas introdutórias que tinham como objetivo mapear algumas informações dos perfis dos participantes e deixá-los mais a vontade para que as perguntas relativas aos tópicos da pesquisa fossem realizadas com os participantes já entendendo a mecânica utilizada no restante da entrevista.

Após a fase das perguntas introdutórias, os primeiros questionamentos feitos aos participantes foram se eles conseguiram compreender como a abordagem de montagem dos perfis foi desenvolvida, se eles teriam alguma outra sugestão de como esses dados poderiam ser coletados e trabalhados para que os resultados fossem aprimorados e se conseguiam imaginar alguma outra informação que poderia ser considerada e explorada na montagem dos perfis dos usuários. Estes questionamentos não foram feitos ao mesmo tempo e sim um a um, tendo como principal objetivo verificar se os participantes realmente haviam compreendido como os perfis foram montados para que ao acessarem o protótipo, pudessem entender quais eram as informações que estavam sendo exibidas para eles e como as mesmas foram organizadas.

O resultado destas perguntas foi que, apesar de quase todos os usuários afirmaram terem entendido como a abordagem da pesquisa foi elaborada, percebeu-se durante o desenrolar das outras perguntas, que alguns não tinham um entendimento claro de como

funcionava a montagem do perfil *trending topic*, sendo necessário explicar durante a entrevista como o mesmo foi montado e somente um usuário citou essa dificuldade explicitamente:

“Entendi quase tudo, só fiquei em dúvida quanto ao perfil de trending topic” (Participante CL)

Em relação à organização dos tópicos, dois usuários citaram possíveis melhorias na abordagem:

“Uma classificação mais refinada de temas... tinham 2 categorias pra mim muito similares esporte e jogos... em uma delas, tinham tweets meus sobre neto baiano, que na verdade podem ser classificados nos 2” (Participante ET)

“Sexo... basicamente 90% da internet é isso... pode estar com entretenimento mas eu faria uma categoria” (Participante FS)

O próximo grupo de perguntas se referia às interfaces do protótipo. Foi questionado aos usuários se eles haviam conseguido trocar de forma simples os perfis, selecionar intervalos de tempo e categorias e também ver *posts* relacionados aos termos e tópicos. O principal objetivo desta parte era tentar descobrir se os participantes realmente tinham utilizado o protótipo e o mais importante, verificar se eles realmente haviam entendido como utilizá-lo. Quaisquer dúvidas que porventura existissem em relação a essa parte eram sanadas durante a entrevista e os participantes tinham todo o tempo necessário para utilizar o protótipo de forma correta até que estivessem aptos a continuar a entrevista.

Todos os usuários acharam a interface de consulta dos perfis para avaliação da abordagem simples e intuitiva, porém alguns usuários citaram algumas melhorias

pontuais que facilitariam ainda mais a consulta de informações. Isso fica claro em algumas respostas:

“Na primeira vez que entrei, tentei clicar nos títulos em vermelho de Texto, Link, etc. achando que poderia ir direto pra lá... Então minha sugestão poderia ser de colocar um link ali talvez”
(Participante CL)

“Sim, foi simples, mas talvez eu carregasse os dados mesmo antes de clicar o botão de ok. Aí se fosse refinando os filtros iria recarregando a tela” (Participante AG)

“Quanto ao protótipo, sei, entendi como funciona, mas a janela de resultados ficando sempre atrás atrapalhou. Uma hora eu fiquei um tempão esperando o resultado sem perceber que estava na janela que ficou atrás do navegador principal. Talvez quando você fizer uma nova pesquisa o pop-up possa ser trazido pra frente de novo” (Participante MA)

5.1.2. Montagem do perfil dinâmico

Alguns usuários se mostraram surpreendidos com informações inferidas de seus perfis, pois ao serem questionados se alguns interesses encontrados que considerassem terem sido inferidos corretamente o surpreenderam, as seguintes respostas foram coletadas:

“Não, mas pensei que falava menos de futebol.” (Participante FR)

“Rafael Marques em nono no geral me surpreendeu. Não imaginava nem que falava dele.” (Participante FS) se referindo ao jogador de futebol que atua pelo seu time.

“Sim, 3G. Não tinha ideia que eu reclamava tanto do 3G... E o ‘Java’ faz sentido, nunca notei que escrevi tanto sobre essa bodega”. (Participante MA)

“Renan Calheiros... Meu interesse na verdade é vê-lo fora do senado”. (Participante RG)

Estes trechos de entrevista levam a crer que a abordagem mostra informações intrínsecas dos usuários que muitas vezes nem eles mesmos haviam percebido que se interessavam.

Em relação à precisão dos tópicos inferidos de seus perfis, a maioria dos usuários citou que, apesar dos tópicos representarem corretamente seus interesses, alguns desses mapeamentos de *posts* que deram origem a esses tópicos foram equivocados como pode ser observado nestes depoimentos:

*“Achei coerente, com pequenos desvios como o caso do Davi que saiu como religião” (Participante FR), onde Davi era o nome do filho do entrevistado e um de seus *posts* citava o nome do filho e o protótipo o exibiu no tópico “religião” por considerar um nome bíblico.*

“Official 2012 Honda CR-V Game Day Commercial – ‘Matthew’s Day Off’ Extended Version <http://t.co/gZBzyeET>” (Participante FS) foi categorizado como tópico “jogo” provavelmente em função do termo “Game”, mas faria mais sentido estar categorizado como entretenimento.

“Na verdade eu achei dois esquisitos, mas entendo como foram aferidos. Um foi religião, mas em tweets que eu falo do ‘Santos’, ‘Freiras’ invadem concentração do Flamengo, ‘Sheik’... Mas

respondendo diretamente, acho que eles (fora religião) são um bom resumo dos meus interesses”. (Participante AG)

“Achei a inferência muito boa no sentido de determinação dos grupos de temas. Porém, ela ainda deve ser aprimorada para não colocar dentro dos grupos tweets que sejam flagrantemente não associados àquele grupo. Por exemplo, no grupo ciência e tecnologia há o seguinte tweet: ‘@SalimMaluf Dr. Paulo, atenção ao Mollusco, pois daqui a pouco ele comparará o sr. ao Ali Babá ou, ainda pior, ao Zé Dirceu! aceite isso!!’ Este aqui deveria estar em política.” (Participante JG)

“Não faço idéia, mas achei alguns resultados esquisitos... por exemplo, ele indica 7 posts de religião, mas nenhum é de verdade... dá pra entender em ‘Wando disse que estava na ‘oficina de Deus consertando a turbina’. Deus mandou avisar que deu PT.’, mas não é religioso, logo o filtro precisa ser aprimorado.

Outro exemplo: ‘Ah, moleque! RT @ederign I hope you're ready. Great times are coming. <http://t.co/1ksBWwYu> @leguimas @flsusp @feroult @billcoutinho’ pegou ‘hope’ e achou que ela religião.” (Participante MA)

5.1.3. Extensão da abordagem para outras redes sociais

Ao serem questionados se acreditavam que a abordagem adotada na pesquisa poderia ser aplicada em qualquer rede social baseada primordialmente em mensagens textuais, 90% acreditam que a mesma poderia ser aplicada, sendo, inclusive, citados

alguns exemplos de quais seriam essas redes sociais. Algumas respostas ilustram esta análise:

“Acredito que possa ser até mais preciso. O twitter é um ótimo estudo, mas limitado pelo espaço. Em outras redes talvez possa conseguir muito mais informações ou no mínimo um universo maior para estudar. Eu acho que o LinkedIn e o Facebook seriam bons estudos.” (Participante CL)

“Creio que funcionaria bem com as redes sociais que participo. Acho que os resultados seriam semelhantes. Facebook, talvez o Google Plus.” (Participante FR)

“Excelente! quem gerenciar o DPM terá uma ideia real de quem e como utiliza as redes sociais quase em ‘tempo real’. No pinterest isso seria perfeito, até as categorias e é uma rede que está crescendo.” (Participante FS)

Em relação à melhora da representatividade dos perfis caso a abordagem fosse aplicada em uma rede social que não possuísse limitação de caracteres para postar informações (caso do Twitter), alguns usuários citaram que acreditam que melhoraria, porém outros ficaram reticentes quanto a essa melhora, pois acreditam que poderia causar mais confusão na análise em função da maior quantidade de informações a serem processadas. Alguns depoimentos dos participantes em relação a este tema:

“Com certeza. Por não limitar o conteúdo, a pessoa pode se expressar com mais liberdade sem ter que cortar palavras e escolher as menores. No twitter essa limitação acaba levando a pessoa a abreviações e frases cortadas que podem levar a uma margem de erro maior.” (Participante CL)

“Depende de como é baseado o DPM, mas acho que se for com as mesmas métricas utilizadas no twitter serão mais representativos sim. Um número maior de informações em redes mais utilizadas como o facebook são sempre mais representativos. A maioria dos usuários não tem noção do que falam, curtem, postam ou comentam nas redes sociais.” (Participante FS)

“No Facebook, acho que, por conta do volume, seria mais frutífero. Acho que sem a limitação, a amostra de palavras aumento e fica mais fácil o encaixe dentro de uma categoria. Além disso, em redes como o Facebook, você poderá extrair também os comentários e gerar um perfil mais refinado.” (Participante AG)

“Acredito que sim. Acho quase certo que a maior variedade de vocábulos e a menor necessidade de abreviações tornarão o mecanismo de inferência muito mais preciso.” (Participante JG)

“Poderia ser bem diferente pelo motivo que citei: se a classificação dos tweets não leva em consideração uma combinação de palavras, quanto mais palavras (falta de limite de caracteres) maior a chance de ser impreciso na classificação.” (Participante ET)

5.2. Análise dos objetivos traçados

Dois objetivos foram inicialmente traçados na pesquisa e após a execução das entrevistas, foi possível realizar uma análise para verificar se os mesmos foram atingidos.

O objetivo principal foi capturar dinamicamente o perfil dos usuários verificando se os mesmos acreditam que foi gerado um perfil representativo deles na rede social objeto do estudo. Já o objetivo secundário foi verificar entre os perfis existentes, qual foi o mais representativo dos interesses dos usuários.

Uma das perguntas do questionário foi se dentre os tipos de perfis apresentados, algum conseguiu mapear os perfis dos participantes de forma satisfatória e, em caso positivo, foi questionado qual seria esse perfil. Continuando a pergunta, foi averiguado ainda se apenas um ou mais de um dos tipos perfis existentes (texto, *link* e *trending topic*) conseguiu esse mapeamento, pedindo, inclusive, para citá-los. O resultado é apresentado na Figura 17.

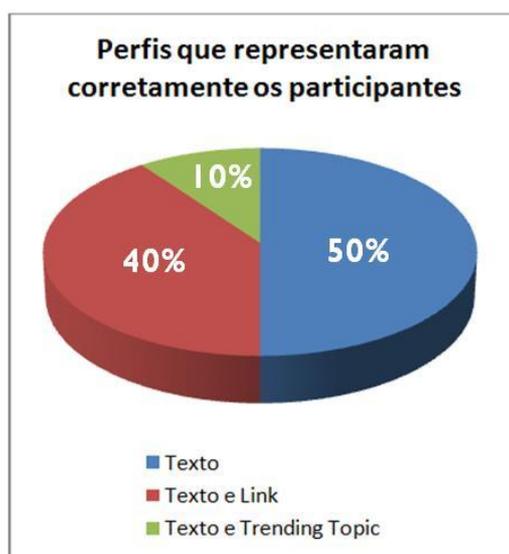


Figura 17 – Tipos de perfis apontados como corretos pelos participantes

Para reforçar a questão da representatividade dos perfis, foi questionado se os interesses (representado as entidades nomeadas) trazidos na pesquisa do seu(s) perfil(s) mais representativo(s) realmente condizem com os interesses dos participantes e se haviam sentido falta de algum interesse. Eis alguns depoimentos:

“Eu diria que 80% sim. Na verdade não (sentiu falta de algum perfil), mas o 80% eu dei por ter coisas que eu na verdade retwitto e não considero exatamente ‘meu perfil’.” (Participante CL),

sendo importante fazer um adendo que os *retweets* (*posts* de outros usuários que são referenciados pelo participante) citados, realmente foram considerados nos perfis construídos dos participantes, pois apesar de ser uma menção a outros usuários, a pesquisa considera que o seu conteúdo é de fato do interesse do participante que realizou o *retweet*.

“Sim, de maneira geral, sim. Apenas quando aparece, por exemplo, Java, mas eu não tenho nenhum interesse em java, mas sim em tecnologia ou desenvolvimento.” (Participante MA)

“Acho que condizem com os interesses baseados nos Tweets.” (Participante RG)

“Sim. Acho que está bem classificado.” (Participante WS)

Foi solicitado também aos participantes que realizassem a ordenação dos perfis, do mais representativo para o menos representativo e a Figura 18 apresenta os resultados.

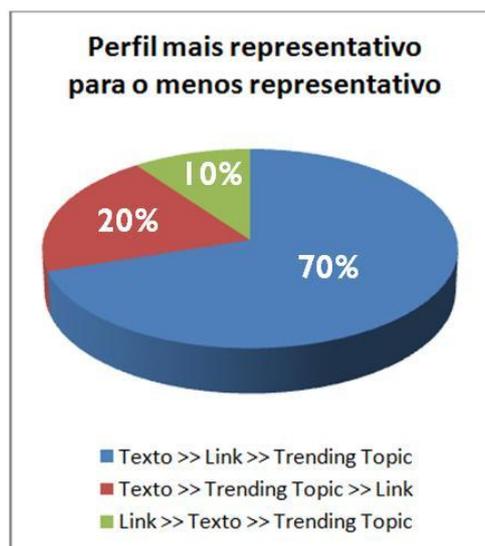


Figura 18 – Ordenação pela representatividade dos tipos de perfis

Houve consenso de que pelo menos um dos perfis conseguiu realizar o mapeamento de forma correta, tendo alguns usuários citados ainda que mais de um tipo

de perfil o representou corretamente, conforme mostra a Figura 17.

Analisando os resultados é possível perceber que o perfil texto foi o que melhor representou os usuários, pois em 90% dos casos foi o que ficou em primeiro lugar na lista de perfis mais representativos, além de ter sido o único que em 100% dos casos representou o perfil dos participantes de forma correta.

Foi possível perceber também que o perfil *trending topic* foi o mais apontado como sendo o menos representativo por 80% dos participantes, além de ter representado corretamente o perfil de somente 10% dos entrevistados.

Voltando aos objetivos e analisando os dados obtidos das entrevistas, os resultados apontam que o objetivo principal foi alcançado, pois todos os usuários citaram que ao menos um dos perfis os representou corretamente e o objetivo secundário tem seu resultado representado na Figura 18.

5.3. Ponto de saturação

Em relação ao tamanho da amostra a ser utilizada, o MEDS cita que a adequação do emprego de amostras pequenas nas pesquisas qualitativas é consensual, onde o número de participantes raramente é estipulado inicialmente, pois o principal critério usado para determinar se as entrevistas realizadas são suficientes para a investigação de um determinado assunto é o da saturação da informação (NICOLACI-DA-COSTA, 2009).

Apesar de 15 usuários terem sido selecionados para participarem do estudo de caso, após a análise das 10 entrevistas iniciais, concluiu-se que não haveria mais informações novas ou relevantes que pudessem ser consideradas na pesquisa e por isso acredita-se que o ponto de saturação citado foi atingido.

Capítulo 6 – Conclusão

Neste capítulo serão apresentadas as conclusões finais desta pesquisa baseando-se primordialmente nos resultados obtidos a partir do estudo de caso realizado, assim como suas contribuições e limitações. Por último, serão apresentadas algumas sugestões de trabalhos futuros para expansão da pesquisa realizada.

6.1. Conclusões

Segundo os resultados obtidos na avaliação, a abordagem proposta se mostrou válida para mapear perfis de usuários dinamicamente através de análise de mensagens textuais que são disponibilizadas de forma pública na rede social objeto de estudo.

O perfil texto, segundo 90% dos usuários, foi o que melhor os representou, sendo também o único que em 100% dos casos representou o perfil dos participantes de forma correta. Isso é algo muito interessante, pois mesmo se tratando de uma rede social onde existe uma limitação máxima de 140 caracteres para inserção de informações nos *posts*, sendo que ainda muitas dessas informações podem ser *links* ou *hashtags*, os resultados apontam que as entidades nomeadas encontradas somente no texto desses *posts*, assim como os tópicos inferidos são os que melhor representaram o perfil desses usuários. Levando em consideração outras redes sociais que não possuam esse tipo de limitação de caracteres, a tendência é que os perfis inferidos sejam ainda mais representativos,

algo que os usuários objeto da pesquisa também acreditam, mas para isso é necessário que a abordagem seja estendida a outras redes sociais que possuam os elementos necessários para análise, como o Google Plus e o Facebook.

O perfil *link* foi o segundo escolhido pelos usuários. A tarefa de processar *links* sem ter uma noção exata de qual conteúdo será encontrado nos mesmos ou até tratar estruturas de páginas com códigos HTML bem distintos é algo bastante complexo, mas mesmo assim os resultados obtidos levam a crer que esse tipo de informação não pode ser descartada. ABEL *et al.* (2011) em seu trabalho também realizam análise de conteúdos presentes em URLs, mas limitam sua atuação somente em páginas de artigos de algumas instituições americanas, algo mais simples de ser tratado.

O perfil *trending topic* foi o que obteve menos sucesso representando o perfil dos participantes. Isso pode ser explicado por uma limitação da API de busca do Twitter que não permite buscar *posts* de uma determinada data a não ser que esta data esteja compreendida entre 6 e 9 dias antes da data atual. A ideia inicial era buscar o contexto exato da *hashtag* através da data de criação do *post* para que fosse possível buscar os termos relevantes e adicionar no perfil dos usuários, mas essa limitação de tempo faz com que não se consiga buscar o dia exato em que a *hashtag* foi mencionada, pois muitas vezes uma *hashtag* é algo que só faz sentido durante um determinado intervalo de tempo, como, por exemplo, algum evento ou show que está ocorrendo naquele momento.

Tentou-se minimizar essa situação trazendo os *posts* mais “populares” de cada *hashtag*, mas mesmo assim, devido às métricas de recuperação dessas informações do Twitter, muitas vezes o texto que era retornado não fazia muito sentido em relação à *hashtag* passada. Importante ressaltar não foi encontrada uma explicação do que significa o valor “popular” utilizado na API de busca do Twitter que traz os *posts* mais

populares para determinada *hashtag*. Acredita-se que sejam *posts* muito mencionados por outros usuários.

6.2. Contribuições

As contribuições desta pesquisa são as seguintes:

- A forma com que os perfis foram construídos e as informações foram buscadas não obriga os usuários a ficarem atualizando forçosamente seus perfis estáticos e faz com que seus perfis estejam sempre atualizados, pois a abordagem busca informações que os usuários inserem de forma espontânea e as torna explícita facilitando assim sua interoperabilidade;
- Através desta abordagem as informações inferidas nos perfis dos usuários poderão ser utilizadas para melhorar a experiência dos próprios usuários quando estiverem utilizando essas redes sociais. Talvez utilizar essas informações inferidas para outra rede social onde este usuário também possua acesso ou até para fins de recomendação de produtos e serviços, seriam exemplos válidos de utilização dos dados dos perfis construídos por esta abordagem;
- Acredita-se também que a arquitetura desenvolvida poderá ser aplicada de forma genérica para outras redes sociais, pois muitas delas também possuem os principais elementos trabalhados nas mensagens inseridas de cada usuário que são: o próprio texto, as URLs que apontam para outras páginas web e marcações específicas para categorizar mensagens que são as *hashtags*;
- Explicitação de características intrínsecas dos usuários que antes de utilizarem o DPM não tinham conhecimento que se interessavam

(positivamente ou negativamente) por certos assuntos inferidos em seus perfis;

- Como uma contribuição técnica é possível citar a construção do protótipo e todo o código-fonte do DPM, os quais poderão ser estendidos para outros trabalhos que tenham um foco similar a este.

6.3. Limitações e generalizações

Não é incomum encontrarmos pesquisas relacionadas ao mapeamento de perfis de usuários que se focam bastante na geração de dados quantitativos, considerando que os perfis construídos dos usuários estão corretos, sem ao menos consultá-los (mesmo que em uma pequena amostragem), sendo que muitas vezes o próprio pesquisador realiza essas análises nos perfis dos usuários para checar se suas informações inferidas estão corretas.

Este não foi o foco deste trabalho, ou seja, o que essa pesquisa se propôs a fazer foi construir perfis distintos de usuários considerando diferentes métricas, realizar comparações destes perfis, mas sempre levando em consideração o que o próprio usuário estava informando sobre seus perfis construídos. Por isso a grande importância de se utilizar o MEDS como ferramenta, pois através dele foi possível explicitar o que os usuários estavam achando da abordagem e dos seus perfis inferidos durante as entrevistas.

As conclusões obtidas na pesquisa podem ser generalizadas mesmo tendo sido realizada uma análise qualitativa, pois a metodologia MEDS adota a posição de que é possível fazer generalizações a partir de amostras intencionais homogêneas de pequeno porte e seus resultados podem ser generalizados para os membros da população mais

ampla, unicamente quando estes exibem as mesmas características que serviram de base para o recrutamento dos participantes da pesquisa (NICOLACI-DA-COSTA, 2007).

Como a pesquisa teve como foco usuários brasileiros, o idioma principal considerado foi o português, porém como não é incomum referências a URLs e *hashtags* em língua inglesa, a mesma também foi tratada. Apesar de somente dois idiomas terem sido utilizados na pesquisa, a solução proposta é flexível o suficiente para qualquer idioma que se deseja utilizar.

Outra limitação encontrada na pesquisa foi citada na seção 6.1, explicando que a API de busca do Twitter¹⁶ não permite buscar *posts* de uma determinada data a não ser que esta data esteja compreendida entre 6 e 9 dias antes da data atual o que provavelmente afetou de forma significativa a construção do perfil *trending topic*.

6.4. Trabalhos futuros

Como trabalho futuro, seria importante melhorar a forma de como o perfil *trending topic* é construído tentando buscar, por exemplo, *posts* de usuários que se relacionam com o usuário sendo pesquisado que contenham esta mesma *hashtag*.

Isso poderia dar uma noção melhor de contexto da *hashtag* do que considerar apenas *posts* de qualquer usuário e não deixaria a aplicação dependente de limitações como foi citado no início desta seção.

Outra ideia para melhorar a construção deste perfil *trending topic* seria utilizar uma abordagem similar à utilizada em (LANIADO & MIKA, 2010) que considera algumas métricas para tentar descobrir se uma *hashtag* serve como um “identificador forte”. Com isso, ela se tornaria relevante o suficiente para ser processada e seus termos extraídos serem considerados no perfil dos usuários, pois algumas *hashtags* na verdade não possuem um contexto que valha a pena ser considerado na montagem dos perfis.

Alguns exemplos deste tipo de *hashtag* que foram encontradas nos *posts* de alguns usuários: #not, #fail, #now.

Trabalhos futuros também envolvem a aplicação da solução proposta em outras redes sociais e também a utilização de uma ontologia que poderia ser utilizada para inferir mais informações dos usuários baseado na similaridade de perfis e também ajudaria na resolução de problemas semânticos como os citados nas análises inter e intra-participantes.

Outra aplicação para utilização de ontologias seria utilizar uma ontologia de domínio para que as informações deste domínio em questão sejam formalmente descritas (ou descrita sem ambiguidades) e com isso, baseado nas informações do perfil dinâmico dos usuários, inferir se determinado usuário tem algum interesse ou seu perfil se assemelha àquele domínio.

A questão de considerar a montagem do perfil dinâmico dos usuários também considerando *posts* de usuários que se relacionam de alguma forma com o usuário pesquisado (no caso do Twitter seriam os *replies* ou também conhecidos como RTs), sendo possível inclusive detectar um grupo de usuários ou usuários mais centrais considerando a combinação de métricas de análise de redes sociais na montagem dos perfis, foi algo sugerido em algumas entrevistas como ilustra o trecho abaixo:

“Talvez fazendo também um consolidado dos seus principais interlocutores.” (Participante AG)

Outra opção de trabalho futuro seria a realização das combinações possíveis dos perfis presentes na abordagem atual fazendo, por exemplo, cálculo de pesos distintos para cada novo tipo de perfil e verificar se os novos também representam corretamente o perfil dos usuários. As combinações possíveis seriam “texto + *link*”, “texto + *link* + *trending topic*”, “texto + *trending topic*” e “*link* + *trending topic*”.

Trabalhos que envolvam a exploração de recomendação de produtos ou conteúdos com base nos perfis construídos dos usuários, também seria uma forma interessante de utilizar as informações obtidas nesta pesquisa.

Por último, talvez uma extensão da abordagem para considerar análise de sentimento em cada entidade nomeada seria algo útil, pois a abordagem atual monta o perfil dos usuários através dos termos que o usuário mais mostrou interesse, mas esse interesse nem sempre é positivo, ou seja, algo que algum usuário não concorde ou o incomode também é citado em seus *posts* e, quando muito citado, aparece no perfil deste usuário. Dois exemplos deste tipo de situação apareceram durante as entrevistas quando os usuários relataram o seguinte:

“E o ‘Java’ faz sentido, nunca notei que escrevi tanto sobre essa bodega.” (Participante MA)

“Renan Calheiros... Meu interesse na verdade é vê-lo fora do senado.” (Participante RG)

Referências

AAKER, D. A.; KUMAR, V.; DAY, G. S. “Marketing Research”, *New York: John Wiley & Sons*, 5th ed, 1995.

ABEL, F.; GAO, Q.; HOUBEN, G.; Tao, K. “Analyzing user modeling on twitter for personalized news recommendations”, *Proceedings of the 19th international conference on User modeling, adaption, and personalization*, July 11-15, Girona, Spain, 2011.

BELKIN, N.J.; CROFT, W.B. “Information filtering and information retrieval: two sides of the same coin?”, *Communications of the ACM - Special issue on information filtering*, Vol. 35, No. 12, pp. 29-38, 1992.

BENBASAT, I., GOLDSTEIN, D.K. and MEAD, M. “The Case Research Strategy in Studies of Information Systems”, *MIS Quarterly*, pp. 369-386, 1987.

BERNERS-LEE, T.; HENDLER, J.; LASSILA, O. “The Semantic Web”, *Scientific American*, pp. 29-37, 2001.

Business Insider, “The state of the Internet”, disponível em <http://www.businessinsider.com/state-of-internet-slides-2012-10?op=1>, 2012, acessado em outubro de 2012.

CARBONELL, J. R. “AI in CAI: an artificial intelligence approach to computer-assisted instruction” *IEEE Transactions on Man-Machine Systems*, Vol. 11, No. 4, pp. 190-202, 1970.

CARR, B.; GOLDSTEIN, I. “Overlays: A Theory of Modelling for Computer Aided Instruction”, *Massachusetts Institute of Technology*, Cambridge AI Lab, 1977.

CRYSTAL, D. “Language and the internet”, *Cambridge University Press*, Cambridge, MA, 2006.

CUMMINS, R. “The Evolution and Analysis of Term-Weighting Schemes in Information Retrieval”, *National University of Ireland*, Galway, Ireland, 2008.

DE SOUZA, C. S.; LEITE, J. C.; PRATES, R. O.; BARBOSA, S. “Projeto de Interfaces de Usuário: Perspectivas Cognitiva e Semiótica”, *Anais da Jornada de Atualização em Informática, XIX Congresso da Sociedade Brasileira de Computação*, Rio de Janeiro, Brasil, 1999.

FIELDING, R. T. “Architectural Styles and the Design of Networked-based Software Architectures”, *Ph.D. dissertation, Information and Computer Science, University of California-Irvine*, Irvine, CA, United States of America, 2000.

FISCHER, G. “User Modeling in Human-Computer Interaction”, *User Modeling and User-Adapted Interaction*, v.11 n.1-2, pp. 65-86, 2001.

GAIZAUSKAS, R.; WILKS, Y. “Information extraction: beyond document retrieval”, *In: Journal of Documentation*, Vol. 54 Iss: 1, pp.70 – 105, 1998.

GIL, A. C. “Como elaborar projetos de pesquisa”, São Paulo: Atlas, 2005.

GODOY, A. S. “Introdução à pesquisa qualitativa e suas possibilidades”, *In: Revista de Administração de Empresas*, v.35, n.2, p. 57-63, 1995.

GRISHMAN, R.; SUNDHEIM, B. “Message Understanding Conference - 6: A Brief History”, *Proceedings of 16th International Conference on Computational Linguistics*, Copenhagen, Denmark, 1996.

HOTH, J.; HALL, W. “An Evaluation of Adapted Hypermedia Techniques Using Static User Modelling”, *Proceedings of the 2nd Workshop on Adaptive Hypertext and Hypermedia*, Pittsburgh, USA, June 20-24, 1998.

KOVAČIČ, T. “Evaluating Web Content Extraction Algorithms”, EngD thesis, *Faculty of Computer and Information Science, University of Ljubljana*, Ljubljana, Slovenia, 2012.

LANIADO, D.; MIKA P. “Making sense of twitter”, *Proceedings of the 9th International Semantic Web Conference on The Semantic Web*, Shanghai, China, 2010.

LIMA, B.; CAMPOS, M. “Recomendação de dados abertos para solucionar os problemas de comunicação textual: Uma análise de métodos para extração de entidades nomeadas”, *In: Brazilian Workshop on Social Network Analysis and Mining (BRASNAM), XXXII Congresso da Sociedade Brasileira de Computação (CSBC)*, Curitiba, Paraná, Brasil, Julho 16-19, 2012.

LUHN, H. “The automatic creation of literature abstracts”, *IBM Journal of Research and Development*, pp 159-165, 1958.

KASS, R.; FININ, T. “Modelling the user in natural language systems”, *Computational Linguistics*, Vol. 14, No. 3, pp. 5-22, 1988.

KINNEAR, T. C.; TAYLOR, J. R. “Marketing research: An applied approach”, *New York: McGraw-Hill Inc.*, 5th Ed., 1996.

MISLOVE, A.; VISWANATH, B.; GUMMADI, K.; DRUSCHEL, P. “You are who you know: Inferring user profiles in online social networks”, *Proceedings of the 3rd ACM International Conference of Web Search and Data Mining*, February, New York, NY, pp. 251-260, 2010.

MCNEE, S. M.; KAPOOR, N.; KONSTAN, J. A. “Don’t look stupid: avoiding pitfalls when recommending research papers”, *Proceedings of the International Conference on Computer Supported Cooperative Work*, Banff, Alberta, Canada, pp. 171-180, 2006.

MOENS, M. F., “Information Extraction: Algorithms and Prospects in a Retrieval Context”, *The Information Retrieval Series*. Vol. 21. 1 ed, Softcover, 2006.

NADEAU, D.; SEKINE, S. “A survey of named entity recognition and classification”, *Linguisticae Investigationes*, Vol. 30, No. 1, pp. 3-26, January, 2007.

NEVES, J. L. “Pesquisa qualitativa: características, usos e possibilidades”, *Caderno de Pesquisas em Administração*, São Paulo, v. 1, n. 3, 1996.

NICOLACI-DA-COSTA, A. M., “O campo da pesquisa qualitativa e o Método de Explicitação do Discurso Subjacente (MEDS)”, *Psicologia: Reflexão e Crítica*. Porto Alegre, v. 20, n.1, 2007.

NICOLACI-DA-COSTA, A. M.; ROMÃO-DIAS; D.; LUCCIO, F. D. “O Uso de Entrevistas On-Line no Método de Explicitação do Discurso Subjacente (MEDS)”, *Psicologia. Reflexão e Crítica*, v. 22, p. 36-43, 2009.

NICOLACI-DA-COSTA, A. M.; LEITÃO, C.; ROMÃO-DIAS, D. “Como conhecer usuários através do Método de Explicitação do Discurso Subjacente (MEDS)”, *In: Proceedings of the 6th Symposium on Human Factors in Computer Systems*, Sociedade Brasileira de Computação, Curitiba, p. 49-59, 2004.

NUNES, S.; DAVID G. “Uma Arquitetura Web para Serviços Web”, *In: XATA - XML: Aplicações e Tecnologias Associadas*, 3ª Conferência Nacional, Universidade do Minho, Braga, p.205-215, 2005.

PARASURAMAN, A. “Marketing Research”, *New York: Addison-Wesley Publishing Company*, p.129, 1991.

PERKOWITZ, M.; ETZIONI, O. “Adaptive web sites: an AI challenge”, *Proceedings of the 15th international joint conference on Artificial intelligence*, p.16-21, August 23-29, Nagoya, Japan, 1997.

PRATES, J. C. “Técnicas de segmentação e agrupamento aplicadas a recursos predeterminados para contextualizar buscas na Internet”. Dissertação de M.Sc., *PPGI/UNIRIO*, Rio de Janeiro, RJ, Brasil, 2011.

RUBIN, H. J.; RUBIN, I. S. “Qualitative interviewing: The art of hearing data”, *Thousand Oaks, CA: Sage*, 2nd ed., 2005.

SARAWAGI, S. “Information Extraction”, *Now Publishers Inc*, 2008.

SCARINCI, R. G. “Extração de Informação como Base para Descoberta de Conhecimento em Dados não Estruturados”, *In: I Workshop sobre Descoberta de Conhecimento em Bases de Dados*, Instituto de Informática da UFRGS, Porto Alegre, Brasil, 2000.

SEIDMAN, I. “Interviewing as qualitative research: A guide for researchers in education and the social sciences”, *New York: Teachers College Press*, 2nd ed., 1998.

SELF, J. “Bypassing the intractable problem of student modeling”, *Proceedings of the International Conference on Intelligent Tutoring Systems*, Montreal, Canada, pp. 18-24, 1988.

SOSNOVSKY, S.; DICHEVA, D. “Ontological technologies for user modeling”, *Int. J. Metadata, Semantics and Ontologies*, Vol. 5, No. 1, pp.32-71, 2010.

SUGIYAMA, K.; HATANO, K.; YOSHIKAWA, M. “Adaptive web search based on user profile constructed without any effort from users”, *Proceedings of the 13th international conference on World Wide Web*. May 17-20, New York, NY, USA, 2004.

TAYLOR, S. “Locating and conducting discourse analytic research”, *In: Wetherell, Margaret; Taylor, Stephanie and Yates, Simeon J. eds. Discourse as Data: A Guide for Analysis*, London: Sage Publications Ltd., pp. 5-48, 2001.

TechCrunch, “Twitter May Have 500M+ Users But Only 170M Are Active, 75% On Twitter’s Own Clients”, disponível em <http://techcrunch.com/2012/07/31/twitter-may-have-500m-users-but-only-170m-are-active-75-on-twitters-own-clients>, 2012, acessado em outubro de 2012.

TURATO, E. R., “Tratado da metodologia da pesquisa clínico-qualitativa”, *Petrópolis, RJ: Vozes*, 2003.

WEISS, R. S. “Learning from strangers: The art and method of qualitative interview studies”, *New York: The Free Press*, 1994.

WENG, J.; LIM, E.; JIANG, J.; HE, Q. “TwitterRank: finding topic-sensitive influential Twitterers”, *In Davison, B.D., Suel, T., Craswell, N., Liu, B., eds.: Proc. of 3rd ACM International Conference on Web Search and Data Mining*. New York, NY, USA, ACM 261-270, 2010.

YIN, R. K. “Estudo de Caso: Planejamento e Métodos”, 3ª edição, Porto Alegre, Ed. Bookman, 2005.

Anexo I – Roteiro da Entrevista

Introdução

- Você acessa a internet todos os dias?
- Qual o dispositivo mais utilizado para acessar a internet?
- Com que frequência faz uso de redes sociais? Quais?

Como aprimorar a abordagem e a consulta das informações

- Você conseguiu compreender como a abordagem de montagem do perfil foi desenvolvida? Teria alguma outra sugestão de como esses dados poderiam ser coletados e trabalhados para que os resultados fossem aprimorados? Consegue imaginar alguma outra informação que poderia ser considerada e explorada na montagem dos perfis dos usuários?
- O que você achou da interface do protótipo? Conseguiu trocar os perfis, selecionar intervalos de tempo e categorias, ver posts relacionados aos termos e assuntos de forma simples?
- Tem alguma sugestão de recursos que poderiam ser adicionados à interface de consulta? (Em caso positivo, perguntar: Qual(is)? Em que aspecto melhoraria?)

Montagem do perfil dinâmico

- Dentre os tipos de perfis apresentados, você acredita que algum conseguiu mapear de forma satisfatória o seu perfil? (Em caso positivo, perguntar: Qual(is)? Conseguiria ordenar estes perfis do mais representativo para o menos representativo?)
- Os interesses trazidos na pesquisa do seu perfil mais representativo realmente condizem com os seus interesses? Sentiu falta de algum?
- Algum interesse encontrado que você considere ter sido inferido corretamente te surpreendeu? (Em caso positivo, perguntar: Qual(is)?)
- O que você achou dos assuntos que foram inferidos no perfil que você considerou mais representativo? (Caso a resposta não cite se os assuntos são realmente de seu interesse, perguntar: São realmente assuntos que você possui interesse?) Sentiu falta de algum assunto? Lembrando que existia uma gama finita de assuntos. São eles: artes e entretenimento, negócios, computadores e internet, cultura e política, jogos, saúde, temas policiais, religião, lazer, ciência e tecnologia, esportes e previsão do tempo.
- Você considera assertivo o mapeamento dos posts que deram origem aos seus interesses? Achou algum(ns) post(s) que não fazia(m) sentido para um determinado interesse ou assunto? (Em caso positivo, perguntar: Conseguiria citar algum?)

Extensão da abordagem para outras redes sociais

- Como você vê a utilização desta abordagem para montar os perfis em outras redes sociais? (Verificar se a resposta já menciona a extensibilidade, caso contrário, perguntar: Você acredita que a abordagem utilizada para a

montagem dos perfis também seria extensível para outras redes sociais?

(Em caso positivo, perguntar: Poderia citar alguns exemplos?)

- Acredita que em uma rede social a qual não possua limitação de caracteres para postar informações, os perfis inferidos serão mais representativos? Por quê?

Anexo II – Texto das Entrevistas

Todas as entrevistas estão descritas fielmente, contendo, inclusive linguagem coloquial e até alguns pequenos erros gramaticais devido à natureza informal com que as mesmas foram executadas.

As entrevistas foram numeradas de 1 a 10 e contêm inicialmente o codinome do participante, a data que foi executada e o software utilizado.

1. Participante FR entrevistado no dia 25/03/2013 via Microsoft Office Communicator

10:14 Gustavo Fernandes
Vamos às perguntas.
Você acessa a internet todos os dias?

10:14 Participante FR
Sim

10:14 Gustavo Fernandes
Qual o dispositivo mais utilizado para acessar a internet?

10:14 Participante FR
smartphone

10:14 Gustavo Fernandes
Com que frequência faz uso de redes sociais?

10:15 Participante FR
diariamente

10:15 Gustavo Fernandes
Quais?

10:15 Participante FR
facebook e twitter

10:16 Gustavo Fernandes
Em relação à montagem dos perfis, ficou claro como a mesma foi realizada?

10:16 Participante FR
sim

10:20 Gustavo Fernandes
Dentre os tipos de perfis apresentados, qual você considera o que representa melhor o seu perfil?

10:20 Participante FR
Creio que seja o de texto

10:20 Gustavo Fernandes
Por quê?

10:22 Participante FR
Porque considera apenas o conteúdo dos posts. E foi o que o que mais se aproximou das categorias de interesse

10:22 Gustavo Fernandes
Conseguiria ordenar os perfis do mais representativo para o menos representativo?

10:22 Participante FR
Texto
Links
TT

10:22 Gustavo Fernandes
Os interesses trazidos na pesquisa do seu perfil mais representativo condizem com os seus interesses?

10:23 Participante FR
Sim

10:23 Gustavo Fernandes
Sentiu falta de algum?

10:23 Participante FR
Na verdade não. Achei abrangente.

10:24 Gustavo Fernandes
Algum interesse encontrado que você considere ter sido inferido corretamente te surpreendeu?

10:25 Participante FR

Não...mas pensei que falava menos de futebol. rs

10:26 Gustavo Fernandes

O que você achou dos assuntos que foram inferidos no perfil que você considerou mais representativo? Lembrando que os assuntos eram exibidos no gráfico de pizza.

10:27 Participante FR

Se não me engano, "arte e cultura"

10:29 Gustavo Fernandes

Na verdade a pergunta não foi clara. O que preciso saber é se no perfil que você considerou mais representativo, no seu caso o texto, o que você achou dos assuntos inferidos?

10:31 Participante FR

ah...achei coerente, com pequenos desvios como o caso do Davi que saiu como religião. Mas a grosso modo, foi coerente.

10:34 Gustavo Fernandes

Correto, mas neste caso específico do Davi, quando você citou este termo em um de seus posts, qual a conotação deste termo? Ou seja, seria um artista ou jogador, por exemplo?

10:35 Participante FR

Davi, neste caso, é o nome do meu filho.

10:36 Gustavo Fernandes

OK. Ainda em relação aos assuntos exibidos no seu perfil texto, são realmente assuntos que você possui interesse?

10:36 Participante FR

sim

10:36 Gustavo Fernandes

Sentiu falta de algum assunto?

10:37 Participante FR

Não. Creio que sejam estes mesmos os assuntos comentados nas redes sociais.

10:37 Gustavo Fernandes

Você considera assertivo o mapeamento dos posts que deram origem aos seus interesses?

10:38 Participante FR

sim

10:38 Gustavo Fernandes

Achou algum(ns) post(s) que não fazia(m) sentido para um determinado interesse ou assunto?

10:40 Participante FR

Achei sim. No caso do post do Davi, onde eu reclamo de ter acordado cedo e saiu com o assunto "Religião".

10:41 Gustavo Fernandes

O que você achou da interface do protótipo?

10:41 Participante FR

Achei simples e de fácil uso.

10:42 Gustavo Fernandes

Conseguiu trocar os perfis, selecionar intervalos de tempo, ver posts relacionados aos termos e filtrar os assuntos de forma simples?

10:42 Participante FR

sim.

10:42 Gustavo Fernandes

Tem alguma sugestão de recursos que poderiam ser adicionados à interface de consulta?

10:44 Participante FR

Poderia, no caso dos links e dos TTs, deixar mais claro o texto usado no processamento.

10:46 Gustavo Fernandes

Deixa ver se entendi. Essa sua sugestão melhoraria o entendimento do porquê de um determinado termo ter sido extraído não do texto do post em si, mas do link ou do trending topic referido por este post?

10:47 Participante FR

isso

10:47 Gustavo Fernandes

OK.

Você conseguiu compreender como a abordagem de montagem do perfil foi desenvolvida?

10:48 Participante FR

sim

10:48 Gustavo Fernandes

Teria alguma outra sugestão de como esses dados poderiam ser coletados e trabalhados para que os resultados fossem aprimorados?

10:53 Participante FR

Creio que a forma como está feita é próxima do ideal. Os desvios são pequenos e não fazem tanta diferença no resultado final.

10:53 Gustavo Fernandes
Consegue imaginar alguma outra informação que poderia ser considerada e explorada na montagem dos perfis dos usuários?

10:54 Participante FR
Sinceramente, não.

10:54 Gustavo Fernandes
Como você vê a utilização desta abordagem para montar os perfis em outras redes sociais?

10:55 Participante FR
Creio que funcionaria bem com as redes sociais que perticipo. Acho que os resultados seriam semelhantes.

10:57 Gustavo Fernandes
Conseguiria citar algum(ns) exemplo(s) de rede(s) social(is) onde esta abordagem se aplicaria?

10:58 Participante FR
Facebook, talvez a Google Plus...

10:58 Gustavo Fernandes
Acredita que em uma rede social a qual não possua limitação de caracteres para postar informações, os perfis inferidos serão mais representativos?

10:59 Participante FR
Creio que sim, mas talvez os desvios sejam maiores também.

11:02 Gustavo Fernandes
Terminaremos por aqui. Obrigado pela sua participação.

2. Participante AG entrevistado no dia 01/04/2013 via Facebook Chat

2:43pm - Gustavo Fernandes
Você acessa a internet todos os dias?

2:44pm - Participante AG
sim

2:44pm - Gustavo Fernandes
Qual o dispositivo mais utilizado para acessar a internet?

2:45pm - Participante AG
bem, na pratica é o desktop por que eu fico conectado o dia inteiro, mas acho que se contarmos cada entrada no celular com +1 acesso eu acho que seria o celular.

2:45pm - Gustavo Fernandes
Com que frequência faz uso de redes sociais?

2:46pm - Participante AG
diariamente

2:46pm - Gustavo Fernandes
Quais?

2:47pm - Participante AG
facebook, instagram e twitter todos os dias. LinkedIn uma ou duas vezes na semana. As outras eu abandonei

2:47pm - Gustavo Fernandes
Você conseguiu compreender como a abordagem de montagem do perfil foi desenvolvida?

2:49pm - Participante AG
me parece que foi por semelhança de palavras utilizadas, palavras que, de algum modo estão categorizadas.

2:49pm - Gustavo Fernandes
Mas conseguiu compreender os três tipos distintos de perfis que foram montados?

2:49pm - Participante AG
sim.

2:50pm - Gustavo Fernandes
Teria alguma outra sugestão de como esses dados poderiam ser coletados e trabalhados para que os resultados fossem aprimorados?

2:51pm - Participante AG
Talvez fazendo também um consolidado dos seus principais interlocutores. Tudo bem que eles não são determinantes de um comportamento, mas talvez sejam se os mesmos foram perfis corporativos...

2:52pm - Gustavo Fernandes
Entendi. Na verdade seria considerar as pessoas com as quais você se comunica e dá "RT" em mensagens, correto?

2:53pm - Participante AG
isso

2:53pm - Gustavo Fernandes
OK. Consegue imaginar alguma outra informação que poderia ser considerada e explorada na montagem dos perfis dos usuários?

2:56pm - Participante AG
sei lá... falando sem pensar muito... integrações entre redes sociais... algo do tipo "vc compartilhou 20% de seus tweets no facebook, outros 30 no linkedin" e ver que outros dados da pra extrair dessas integrações... sei lá.... tem que dar uma pensada

2:56pm - Gustavo Fernandes

OK. O que você achou da interface do protótipo? Conseguiu trocar os perfis, selecionar intervalos de tempo e categorias, ver posts relacionados aos termos e assuntos de forma simples?

2:58pm - Participante AG

Sim, foi simples, mas talvez eu carregasse os dados mesmo antes de clicar o botão de ok. Aí se fosse refinando os filtros iria recarregando a tela...

2:59pm - Gustavo Fernandes

OK. Dentre os tipos de perfis apresentados, você acredita que algum conseguiu mapear de forma satisfatória o seu perfil?

3:01pm - Participante AG

usei bem mais o de Texto e confesso que me surpreendi com o volume de algumas categorias.

3:02pm - Gustavo Fernandes

Então você considera que o perfil texto foi o que melhor representou o seu perfil, correto?

3:02pm - Participante AG

sim

3:03pm - Gustavo Fernandes

Além dele, mais algum você considera que tenha representado o seu perfil satisfatoriamente? Link ou Trending Topic?

3:05pm - Participante AG

acho que não, mas provavelmente pelo volume menor.

3:05pm - Gustavo Fernandes

Conseguiria ordenar estes perfis do mais representativo para o menos representativo?

3:06pm - Participante AG

Texto > Link > TT

3:06pm - Gustavo Fernandes

OK. Os interesses trazidos na pesquisa do seu perfil mais representativo realmente condizem com os seus interesses? No seu caso o perfil mais representativo foi o TEXTO.

3:15pm - Participante AG

Ai depende, vejamos. Na lista de termos, descobri que 9/10 principais foram sobre futebol. Apesar de eu gostar, achei exagerado. No gráfico entretanto, achei estranho que somente 2% eram sobre Esporte. Não sei se é um efeito a la long tail (tipo, 80% diluídos entre vários interesses e 20% concentrados em um só, mas vai saber)

3:15pm - Gustavo Fernandes
Entendi. Ainda sobre os interesses, sentiu falta de algum?

3:18pm - Participante AG
assim de primeira, acho que nao

3:18pm - Gustavo Fernandes
Algum interesse encontrado que você considere ter sido inferido corretamente te surpreendeu?

3:20pm - Participante AG
hum... surpreender mesmo, nao.

3:21pm - Gustavo Fernandes
Em uma das respostas anteriores você chegou a comentar sobre os assuntos. O que você achou dos assuntos que foram inferidos no perfil que você considerou mais representativo que foi o TEXTO?

3:23pm - Participante AG
na verdade eu achei dois esquisitos mas entendo como foram aferidos. Um foi religiao, mas em tweets que eu falo do "Santos", "Freiras" invadem concetnracao do Flamengo, "Sheik"... Mas respondendo diretamente, acho que eles (fora religiao) sao um bom resumodos meus interesses

3:25pm - Gustavo Fernandes
Entendi. Um dos usuários que já participou dessa entrevista também citou o tópico religião, pois o nome do seu filho é Davi, ele citou o nome do filho e apareceu no tópico religião por ser um nome bíblico. mas seguindo...

3:27pm - Participante AG
Pois é... mas acho normal ter uma margem de erro. Por exemplo tem um post meu que eu facó uma citacao ao Cuca. Algo bem curtinho, tipo "E o Cuca, hein? " Caiu em Cultura e Politica. Acho normal, ja que tem a Cuca do sitio do pica pau amarelo.

3:27pm - Gustavo Fernandes
É verdade. Uma observação é que todas essas informações foram inferidas automaticamente, sem ajuda de um "humano". bom, voltando... Sentiu falta de algum assunto?

3:30pm - Participante AG
Lendo assim, nao senti nao. Lendo assim, na lista, quero dizer... Ah, mais um exemplo: "Gol do Flu!!!" Saiu em saude. OU por que "Flu" é gripe em ingles ou por que GOls do Flu sao sempre saudaveis!

3:31pm - Gustavo Fernandes
Fico com a tese da gripe.

3:31pm - Participante AG

:)

3:31pm - Gustavo Fernandes

Você considera assertivo o mapeamento dos posts que deram origem aos seus interesses?

3:32pm - Participante AG

Sim. achei bastante válido.

3:32pm - Gustavo Fernandes

Achou algum(ns) post(s) que não fazia(m) sentido para um determinado interesse ou assunto? Na verdade você já até citou o exemplo do Flu. Teria mais algum?

3:34pm - Participante AG

hum, aqueles 3 sobre religiao. deixa eu ver aqui que eu acho que tem outro.

3:34pm - Gustavo Fernandes

Acho que não precisa. Já deu pra ter uma boa idéia. Como você vê a utilização desta abordagem para montar os perfis em outras redes sociais?

3:35pm - Participante AG

blz. acho interessante, mas tem umas distorcoes por conta das proprias redes sociais... Acho que no linkedin, isso daria a impressao que todos sao muito corporativos e focados na sua propria industria. No Facebook, acho que , por conta do volume, seria mais frutifero.

3:39pm - Gustavo Fernandes

Acredita que em uma rede social a qual não possua limitação de caracteres para postar informações, os perfis inferidos serão mais representativos?

3:39pm - Participante AG

sim

3:40pm - Gustavo Fernandes

Conseguiria explicar o por quê?

3:43pm - Participante AG

acho que sem a limitacao, a amostra de palavras aumento e fica mais facil o encaixe dentro de uma categoria. Alem disso em redes como o Facebook vc podera extrair tambem os comentarios e gerar um perfil mais refinado.

3:44pm - Gustavo Fernandes

OK. Acho que é isso. Muito agradecido pelo seu tempo e paciência!

3. Participante MA entrevistado no dia 01/04/2013 via Facebook Chat

7:26pm - Gustavo Fernandes

Você acessa a internet todos os dias?

7:28pm - Participante MA
sim

7:28pm - Gustavo Fernandes
Qual o dispositivo mais utilizado para acessar a internet?

7:28pm - Participante MA
notebook
mas pau a pau com celular

7:29pm - Gustavo Fernandes
Com que frequência faz uso de redes sociais?

7:29pm - Participante MA
todo dia

7:29pm - Gustavo Fernandes
Quais?

7:30pm - Participante MA
linkedin, facebook, twitter...

7:31pm - Gustavo Fernandes
OK. Vamos falar agora um pouco sobre as informações que foram inferidas na montagem do seu perfil.

7:31pm - Participante MA
ok

7:32pm - Gustavo Fernandes
Você conseguiu compreender como a abordagem de montagem do perfil foi desenvolvida?

7:33pm - Participante MA
sim, é possível inferir algumas palavras-chave

7:33pm - Gustavo Fernandes
Teria alguma outra sugestão de como esses dados poderiam ser coletados e trabalhados para que os resultados fossem aprimorados?

7:35pm - Participante MA
não faço idéia, mas achei alguns resultados esquisitos... por exemplo, ele indica 7 posts de religião, mas nenhum é de verdade... dá pra entender em "Wando disse que estava na "oficina de Deus consertando a turbina". Deus mandou avisar que deu PT."
mas não é religioso, logo o filtro precisa ser aprimorado

7:36pm - Gustavo Fernandes
entendi.

7:36pm - Participante MA

outro exemplo: Ah, moleque! RT @ederign I hope you're ready. Great times are coming. <http://t.co/1ksBWwYu> @leguimas @flsusp @feroult @billcoutinho pegou "hope" e achou que ela religião

7:37pm - Gustavo Fernandes

Consegue imaginar alguma outra informação que poderia ser considerada e explorada na montagem dos perfis dos usuários?

7:39pm - Participante MA

putz, nao faço idéia... mas percebo que analisando os outros casos também surgem discrepâncias. além da palavra-chave, talvez o link pode ser analisado, talvez seja mais fácil em um universo maior de palavras-chave do que os 140 caracteres do twitter

7:41pm - Gustavo Fernandes

na verdade o perfil link faz justamente isso, ou seja, a aplicação segue cada link e considera apenas termos presentes nos textos desses links. O mesmo serve para o perfil trending topic.

o perfil texto trata apenas do texto inserido em cada post

Agora sobre a interface do protótipo.

O que você achou da interface do protótipo? Conseguiu trocar os perfis, selecionar intervalos de tempo e categorias, ver posts relacionados aos termos e assuntos de forma simples?

7:44pm - Participante MA

analisando agora os trending topics (só tinha visto texto), também vi coisas estranhas... em cultura e entretenimento, por exemplo:

RT @leguimas: Café com código @DextraSistemas hotdeploy utilizando myContainer e JSF #mycontainer #jsf #hotdeploy <http://t.co/yv4nEdVN>

?@DextraSistemas Artigo imperdível: Why Testing Matters in Agile Projects (via @InfoQ) <http://t.co/5IrgfVC2> #qualidade #agile #testes"

06/10/2012 09:56:49

Ninguém pode reclamar de inconstância: o #3G da @VIVOemrede funciona consistentemente mal todos os dias da semana! #vivo #lixo

7:45pm - Participante MA

Quanto ao protótipo, sei, entendi como funciona, mas a janela de resultados ficando sempre atrás atrapalhou. Uma hora eu fiquei um tempao esperando o resultado sem perceber que estava na janela que ficou atrás do navegador principal.

7:46pm - Gustavo Fernandes

Entendi. Realmente o link é aberto sempre na mesma janela pra evitar ficar abrindo popups infinitos.

Fora essa questão do popup, teria alguma outra sugestão de recursos que poderiam ser adicionados à interface de consulta?

7:47pm - Participante MA

sim, mas talvez quando vc fizer uma nova pesquisa o pop-up possa ser trazido pra frente de novo

7:47pm - Gustavo Fernandes

OK. Agora sobre a montagem do perfil. Dentre os tipos de perfis apresentados, você acredita que algum conseguiu mapear de forma satisfatória o seu perfil?

7:49pm - Participante MA

negócios está bem preciso. esportes também, mas é fácil: só tinha 1 post e com a palavra "olimpiada"

7:50pm - Gustavo Fernandes

Mas eu digo entre os três tipos de perfil: TEXTO, LINK ou TRENDING TOPIC
Você acredita que algum deles conseguiu mapear de forma satisfatória o seu perfil de usuário?

7:52pm - Participante MA

tive a impressão que trending topic ficou mais próximo, mas todos eles tem discrepâncias
trending

7:52pm - Gustavo Fernandes

Entre esses três tipos de perfil, conseguiria ordenar estes perfis do mais representativo para o menos representativo?

7:54pm - Participante MA

só teria certeza analisando com mais calma item a item, então é chute: trending, link e texto.

7:54pm - Gustavo Fernandes

Os interesses trazidos na pesquisa do seu perfil mais representativo, no caso o TRENDING TOPIC, realmente condizem com os seus interesses?

7:56pm - Participante MA

sim, mas não na porcentagem apresentada pq os posts estão classificados errados. pelo menos não tem "religião", por exemplo, como em "texto"

7:56pm - Gustavo Fernandes

entendi, mas essa parte de porcentagem seriam os assuntos e não os interesses que aparecem na tabela

7:57pm - Participante MA

opa, isso está confuso... percebeu que eu confundi todos os termos? não é a primeira vez

7:57pm - Gustavo Fernandes

é verdade.

7:58pm - Participante MA

confesso que ignorei a tabela da direita porque não entendi de primeira o que era e foquei nos gráficos. o gráfico tira o foco da tabela, mexer com imagem é mais atrativo

7:58pm - Gustavo Fernandes

pelo que entendi, vc está se focando somente no gráfico e não nos interesses. problema de UX. anotado.

mas voltando, considerando também os interesses, você ainda acredita que o perfil TRENDING TOPIC é o mais representativo ou gostaria de mudar de idéia?

na verdade os interesses são termos que mais apareceram nos textos dos seus posts, links e trending topics e o gráfico tenta ilustrar e consolidar dos assuntos de cada post.

8:02pm - Participante MA

não, nesse caso parecer ser texto, trending topic e link

8:03pm - Gustavo Fernandes

entendi. Anotado. Voltando então, os interesses trazidos na pesquisa do seu perfil mais representativo, no caso o TEXTO, realmente condizem com os seus interesses?

8:06pm - Participante MA

Sim, de maneira geral, sim. Apenas quando aparece por exemplo Java, mas eu não tenho nenhum interesse em java, mas sim em tecnologia ou desenvolvimento.

8:06pm - Gustavo Fernandes

Sentiu falta de algum?

8:09pm - Participante MA

veja bem, interesses não conseguem ser representados por termos... os termos que aparecem ali listados são verdadeiros, mas eles não são meus interesses. ou seja, eu escrevi muitas vezes "Brasil" mas em contextos diferentes e não dá pra dizer que meu interesse amplo é Brasil

eu poderia ter escrito em época de copa e meu interesse seria futebol, e não o Brasil país

8:09pm - Gustavo Fernandes

entendi.

Algum interesse encontrado que você considere ter sido inferido corretamente te surpreendeu?

8:10pm - Participante MA

está confuso o conceito "Interesse", está muito parecido com "Assunto"

A tabela não é de interesses, mas sim de termos mais escritos que não representam interesses diretamente

8:11pm - Gustavo Fernandes

beleza. Mais um ponto pra verificar no protótipo.

Algum interesse ou termo encontrado que você considere ter sido inferido corretamente te surpreendeu?

8:12pm - Participante MA

Sim, 3G

nao tinha idéia que eu reclamava tanto do 3G

8:13pm - Participante MA

E o "Java" faz sentido, nunca notei que escrevi tanto sobre essa bodega

8:14pm

Gustavo Fernandes

Eu iria te perguntar itens sobre os assuntos (gráfico), mas você já comentou anteriormente que viu alguns problemas como o de religião, por exemplo.

8:15pm - Participante MA

sim, estao bem diferentes

8:15pm - Gustavo Fernandes

Você considera assertivo o mapeamento dos posts que deram origem aos seus interesses (termos)?

8:16pm - Participante MA

sim

8:16pm - Gustavo Fernandes

Achou algum(ns) post(s) que não fazia(m) sentido para um determinado interesse ou assunto?

Na verdade você já até disse que sim, correto?

8:17pm - Participante MA

sim, disse

liste inclusive alguns no começo da conversa

8:17pm - Gustavo Fernandes

É verdade.

Como você vê a utilização desta abordagem para montar os perfis em outras redes sociais?

8:19pm - Participante MA

acho que a filtragem precisa melhorar, independente da rede social... apenas se basear em palavras-chave, se é esse o caso, nao funcionou bem... nos assuntos errou muito na classificação e nos interesses nao relacionou a quantidade de termos com interesses de verdade (mas aí pode ser só interpretação da palavra "interesses")

8:20pm - Gustavo Fernandes

Acredita que em uma rede social a qual não possua limitação de caracteres para postar informações, os perfis inferidos serão mais representativos?

8:22pm - Participante MA

em teoria, sim, mas é um chute de um leigo

8:22pm - Gustavo Fernandes

Obrigado pelo seu tempo e paciência. Suas respostas foram muito úteis!

4. Participante SP entrevistado no dia 02/04/2013 via Microsoft Office Communicator

09:53 - Gustavo Fernandes

Você acessa a internet todos os dias?

09:53 - Participante SP

sim!

09:54 - Gustavo Fernandes

Qual o dispositivo mais utilizado para acessar a internet?

09:54 - Participante SP

PC-Desktop

09:54 - Gustavo Fernandes

Com que frequência faz uso de redes sociais?

09:54 - Participante SP

todos os dias

09:55 - Gustavo Fernandes

Quais?

09:55 - Participante SP

Twitter e Facebook

09:55 - Gustavo Fernandes

Vamos falar agora um pouco sobre as informações que foram inferidas na montagem do seu perfil.

Você conseguiu compreender como a abordagem de montagem do perfil foi desenvolvida?

Pergunto sobre a forma de montar os três perfis: TEXTO, LINK e TRENDING TOPIC.

09:56 - Participante SP

mais ou menos

isso sim!

09:56 - Gustavo Fernandes

OK. Teria alguma outra sugestão de como esses dados poderiam ser coletados e trabalhados para que os resultados fossem aprimorados?

09:58 - Participante SP

No Twitter vc consegue enviar mensagem para uma pessoa especifica, apesar do escopo não ser esse

09:59 - Gustavo Fernandes

entendi. Seria considerar também a interação com outros usuários, correto?

09:59 - Participante SP

sim

09:59 - Gustavo Fernandes

OK. Consegue imaginar alguma outra informação que poderia ser considerada e explorada na montagem dos perfis dos usuários?

10:01 - Participante SP

no Facebook existem pessoas que gostam de "expor" a vida social, outras como eu, gostam apenas de ler besteiras e piadas. Entende a utilização de maneira diferente

da mesma forma que no twitter tem gente que só escreve besteira, outros compartilham informações de trânsito, etc são focos diferentes para a mesma aplicação

10:02 - Gustavo Fernandes

OK. O que você achou da interface do protótipo? Conseguiu trocar os perfis, selecionar intervalos de tempo e categorias, ver posts relacionados aos termos e assuntos de forma simples?

10:03 - Participante SP

sim, tranquilo

10:03 - Gustavo Fernandes

Tem alguma sugestão de recursos que poderiam ser adicionados à interface de consulta?

10:04 - Participante SP

nada me vem a mente agora

10:04 - Gustavo Fernandes

Agora vamos falar sobre a montagem do perfil.

Dentre os tipos de perfis apresentados, você acredita que algum conseguiu mapear de forma satisfatória o seu perfil?

Lembrando que os tipos de perfil são: TEXTO, LINK e TRENDING TOPIC.

10:05 - Participante SP

sim, totalmente!

10:06 - Gustavo Fernandes

Qual(is)? Pra facilitar, conseguiria ordenar estes perfis do mais representativo para o menos representativo?

10:07 - Participante SP

Texto

10:08 - Gustavo Fernandes

Depois do texto, qual você acha que mais se assemelhou ao seu perfil de usuário?
LINK ou TRENDING TOPIC?

10:08 - Participante SP

Link

10:09 - Gustavo Fernandes

OK. Os interesses trazidos na pesquisa do seu perfil mais representativo, no caso o TEXTO, realmente condizem com os seus interesses?

10:09 - Participante SP

sim!

10:09 - Gustavo Fernandes

Sentiu falta de algum? Algum interesse?

10:10 - Participante SP

não! Só 1 detalhe esta "enimento"

10:10 - Gustavo Fernandes

na verdade é um problema no gráfico, mas quando você passa o mouse em cima dele, aparece o texto completo.

10:11 - Participante SP

ok

10:11 - Gustavo Fernandes

Algum interesse encontrado que você considere ter sido inferido corretamente te surpreendeu?

10:12 - Participante SP

não, pelo contrário

10:12 - Gustavo Fernandes

O que você achou dos assuntos que foram inferidos no perfil que você considerou mais representativo? Lembrando que os assuntos são exibidos no gráfico.

10:14 - Participante SP

eu falo muito twitter sobre política, então a fatia deveria ser maior no meu entendimento

10:15 - Gustavo Fernandes

OK. Sentiu falta de algum assunto?

10:15 - Participante SP

não!

10:15 - Gustavo Fernandes

Pelo que você respondeu não sentiu falta, mas acredita que política deveria ter tido uma fatia maior, correto?

10:16 - Participante SP

considerando que assuntos pessoais estão em entreterimento, o q não sei se deveria

10:16 - Gustavo Fernandes entendi.

10:17 - Participante SP

talvez uma fatia de mensagens pessoais seria interessante. Por exemplo: A Monise tuitava que estava com sono,rs
eu escrevo mais sobre a minha indignação política

10:17 - Gustavo Fernandes

OK. Dica anotada. Você considera assertivo o mapeamento dos posts que deram origem aos seus interesses (tabela) e assuntos (gráfico)?

10:18 - Participante SP

não! nos meus interesses com o perfil de texto aparece somente "SmartPhones"
não concordo

10:20 - Gustavo Fernandes

Mas isso para uma determinada categoria e em um determinado intervalo de tempo ou sem filtro algum considerando somente a combo box de perfil?

10:21 - Participante SP

ah tá verdade, rs

10:22 - Gustavo Fernandes

Isso é até normal, porque quanto mais filtros você colocar na busca, menos itens vão aparecer pra você na resposta.

10:22 - Participante SP

corrigindo: Esta bem assertiva

10:22 - Gustavo Fernandes

OK. Achou algum(ns) post(s) que não fazia(m) sentido para um determinado interesse ou assunto? Conseguia citar algum exemplo?

10:23 - Participante SP

"Renan Calheiros, Collor, delegado preso toma posse com monitoramento, etc. Alagoas precisa rever seus conceitos"

Esta como entretenimento, eu vejo como política

10:23 - Gustavo Fernandes

Entendi. Como você vê a utilização desta abordagem para montar os perfis em outras redes sociais?

10:24 - Participante SP
muito interessante
vc consegue o comportamento das pessoas

10:24 - Gustavo Fernandes
Você acredita que a abordagem utilizada para a montagem dos perfis também seria extensível para outras redes sociais?

10:24 - Participante SP
com certeza!
Facebook esta no top

10:25 - Gustavo Fernandes
Poderia citar alguns exemplos de redes sociais ainda esta abordagem se aplicaria, além do Facebook que você acabou de citar?

10:28 - Participante SP
não consigo me lembrar, uma vez que as outras são específicas, como a Reddit que são muito específicas

10:28 - Gustavo Fernandes
OK. Acredita que em uma rede social a qual não possua limitação de caracteres para postar informações, os perfis inferidos serão mais representativos?

10:30 - Participante SP
eu sou fã dos 140 caracteres

10:30 - Gustavo Fernandes
Entendi, mas por exemplo, no Facebook que não possui limitação, acredita que os perfis ficariam mais precisos?

10:30 - Participante SP
sim!

10:30 - Gustavo Fernandes
Conseguiria explicar por quê?

10:32 - Participante SP
pode-se explanar a ideia completamente

10:32 - Gustavo Fernandes
Muito agradecido pela sua participação.

5. Participante FS entrevistado no dia 03/04/2013 via Facebook Chat

9:14pm - Gustavo Fernandes
Você acessa a internet todos os dias?

9:15pm - Participante FS

sim, todos os dias

9:15pm - Gustavo Fernandes

Qual o dispositivo mais utilizado para acessar a internet?

9:16pm - Participante FS

notebook

9:16pm - Gustavo Fernandes

Com que frequência faz uso de redes sociais?

9:16pm - Participante FS

tem um padrao? muito, MUITO MUUUUUITO?

9:16pm - Gustavo Fernandes

Quais?

9:17pm - Participante FS

facebook 70% twitter 20%

9:17pm - Gustavo Fernandes

OK. Vamos falar agora um pouco sobre as informações que foram inferidas na montagem do seu perfil.

Você conseguiu compreender como a abordagem de montagem do perfil foi desenvolvida?

9:18pm - Participante FS

consegui

9:18pm - Gustavo Fernandes

Teria alguma outra sugestão de como esses dados poderiam ser coletados e trabalhados para que os resultados fossem aprimorados?

9:19pm - Participante FS

quanto ao sistema sinceramente nao...esta bem redondo, so tenho criticas quanto ao design...

como a base de dados sera maior em breve, o sistema ficara ainda mais redondo algumas respostas ficam um pouco vagas mas pela falta de dados, nao pelo sistema

9:20pm - Gustavo Fernandes

entendi. Consegue imaginar alguma outra informação que poderia ser considerada e explorada na montagem dos perfis dos usuários?

9:21pm - Participante FS

sexo

basicamente 90% da internet é isso

pode estar com entretenimento...mas...eu faria uma categoria

9:23pm - Gustavo Fernandes

entendi, mas digo em relação aos dados que estão disponíveis para coleta nos posts de cada usuário. Consegue imaginar algo a mais que poderia ser considerado para montar os perfis?

9:24pm - Participante FS
nao nao, ta fechado

9:24pm - Gustavo Fernandes
Exemplo: teve um usuário que entrevistei antes que citou que talvez a interação com outros usuários também poderia ser considerada, pois é possível buscar RTs e replies.
OK.

9:24pm - Participante FS
sinceramente nao. acho q ja tem informacao suficiente

9:24pm - Gustavo Fernandes
OK.
O que você achou da interface do protótipo? Conseguiu trocar os perfis, selecionar intervalos de tempo e categorias, ver posts relacionados aos termos e assuntos de forma simples?

9:25pm - Participante FS
sim, ele é bem intuitivo e facil de usar

9:26pm - Gustavo Fernandes
Sua única observação seria realmente em cima do design do protótipo, conforme você comentou antes?

9:26pm - Participante FS
sim, so essa

9:26pm - Gustavo Fernandes
Tem alguma sugestão de recursos que poderiam ser adicionados à interface de consulta?

9:28pm - Participante FS
alguma metrica que eu soubesse sobre oque estao falando de mim. n caso de alguem com muitos seguidores. tipo um pvc da vida. ate pra esse formador de opniao ter como mensurar e seguir um rumo ou uma opniao hunanime ou nao

9:29pm - Gustavo Fernandes
É realmente uma boa idéia. Anotado.
Agora vamos falar sobre a montagem do perfil dinâmico do seu usuário.
Dentre os tipos de perfis apresentados, você acredita que algum conseguiu mapear de forma satisfatória o seu perfil?
Lembrando que os tipos de perfis são: TEXTO, LINK e TRENDING TOPIC.

9:31pm - Participante FS
texto e tranding topic

link pelo menos pra mim ficou um pouco confuso na apresentacao dos resultados mas os outros dois sao bemmmmmm interessantes

9:32pm - Gustavo Fernandes

Conseguiria ordenar estes perfis do mais representativo para o menos representativo?

9:33pm - Participante FS

sim. 1 - REXTO; 2 - TRENDING TOPIC; 3 - LINK

9:33pm - Gustavo Fernandes

OK.

9:33pm - Participante FS

sim. 1 - TEXTO; 2 - TRENDING TOPIC; 3 - LINK

9:33pm - Gustavo Fernandes

Os interesses trazidos na pesquisa do seu perfil mais representativo, no caso o TEXTO, realmente condizem com os seus interesses?

9:34pm - Participante FS

sim, FUTEBOL! RSRRS

9:34pm - Gustavo Fernandes

Sentiu falta de algum?

De algum interesse?

9:35pm - Participante FS

senti...botafogo

9:36pm - Gustavo Fernandes

Algum interesse encontrado que você considere ter sido inferido corretamente te surpreendeu?

9:38pm - Participante FS

Rafael Marques em nono no geral

isso me surpreendeu nao imaginava nem q falava dele

9:38pm - Gustavo Fernandes

Legal (ou não).

9:38pm - Participante FS

hahahaha

9:39pm - Gustavo Fernandes

O que você achou dos assuntos que foram inferidos no perfil que você considerou mais representativo?

Lembrando que os assuntos são exibidos no gráfico.

9:41pm - Participante FS

entretenimento com certeza. Mas cultura e politica foram bem interessantes tb

9:42pm - Gustavo Fernandes

São realmente assuntos que você possui interesse?

9:42pm - Participante FS

sim, muito

so nao achava q era tanto

9:42pm - Gustavo Fernandes

Sentiu falta de algum assunto, além do "sexo" que você citou anteriormente?

9:44pm - Participante FS

nao, basicamente no que pensei se enquadra em algum dos assuntos ja selecionados.

9:44pm - Gustavo Fernandes

Você considera assertivo o mapeamento dos posts que deram origem aos seus interesses?

Lembrando que os posts aparecem no popup depois que um dos interesses é clicado.

9:47pm - Participante FS

a maioria sim. alguns principalmente no perfil LINK nao...

9:48pm - Gustavo Fernandes

Achou algum(ns) post(s) que não fazia(m) sentido para um determinado interesse ou assunto?

9:48pm - Participante FS

sim

9:49pm - Gustavo Fernandes

Conseguiria citar algum?

9:49pm - Participante FS

Official 2012 Honda CR-V Game Day Commercial - "Matthew's Day Off"
Extended Version <http://t.co/gZBzyeET> esta como JOGO

9:50pm - Gustavo Fernandes

Entendi.

9:50pm - Participante FS

seria entretenimento

9:50pm - Gustavo Fernandes

Como você vê a utilização desta abordagem para montar os perfis em outras redes sociais?

9:51pm - Participante FS

excelente! quem gerenciar o DPM tera uma ideia real de quem e como utiliza as redes sociais quase em "tempo real"
ainda mais na parte de TRENDING TOPIC, ele podera observar e "indicar" pra que lado vai o assunto

9:52pm - Gustavo Fernandes

Entendi.

Conseguiria então citar outras redes sociais onde essa abordagem poderia ser aplicada?

9:55pm - Participante FS

o pinterest

isso seria perfeito, ate as categorias e é uma rede q esta crescendo

9:56pm - Gustavo Fernandes

Entendi. Acredita que em uma rede social a qual não possua limitação de caracteres para postar informações, os perfis inferidos serão mais representativos?

9:57pm - Participante FS

depende de como é baseado o DPM. mas acho q se for com as mesmas metricas utilizadas no twitter serao mais representativos sim

9:57pm - Gustavo Fernandes

Conseguiria explicar por que?

9:59pm - Participante FS

um numero maior de informacoes em redes mais utilizadas como o facebook sao sempre mais representativos. a maioria dos usuarios nao tem nocao do que falam, curtem, postam ou comentam nas redes sociais

esse dados para uma profissional na minha area por exemplo nao tem preco (tem na verdade e é bem caro)

10:00pm - Gustavo Fernandes

Entendi.

Chegamos ao fim. Muito agradecido pelo tempo e paciência.

6. Participante CL entrevistado no dia 07/04/2013 via Facebook Chat

7:03pm - Gustavo Fernandes

Você acessa a internet todos os dias?

7:04pm - Participante CL

Sim

7:04pm - Gustavo Fernandes

Qual o dispositivo mais utilizado para acessar a internet?

7:04pm - Participante CL

Computador

7:04pm - Gustavo Fernandes
Com que frequência faz uso de redes sociais?

7:05pm - Participante CL
Todo dia, pelo menos umas três vezes ao dia

7:05pm - Gustavo Fernandes
Quais?

7:06pm - Participante CL
Na ordem de uso: twitter, google+, facebook, linkedin

7:06pm - Gustavo Fernandes
OK.
Vamos falar agora um pouco sobre as informações que foram inferidas na montagem do seu perfil.
Você conseguiu compreender como a abordagem de montagem do perfil foi desenvolvida?

7:07pm - Participante CL
Entendi quase tudo, só fiquei em dúvida quanto ao perfil de trending topic

7:09pm - Gustavo Fernandes
Uma explicação rápido do trending topic então... O protótipo pega a hashtag, faz uma busca no Twitter pelos posts mais relevantes desta hashtag, pega o texto desses posts, processa e extrai os termos.
isso para cada hashtag

7:10pm - Participante CL
Por exemplo, tem lá no meu perfil México, mas abro e vem uma mensagem que escrevi com #EUA. Não entendi bem essa relação...

7:11pm - Gustavo Fernandes
chegou a clicar no ícone "T" que aparece ao lado do post?
ali é mostrado o texto que foi buscado do Twitter aonde esse termo apareceu.
Na verdade isso é só uma tentativa de buscar o contexto da hashtag, mas realmente muitas vezes não faz sentido.

7:12pm - Participante CL
Ou seja, quer dizer que México aparece em mensagens que tem o EUA, que eu postei, seria isso?

7:12pm - Gustavo Fernandes
No caso esse termo México veio como resultado dos posts mais relevantes no Twitter para a hashtag #EUA.
Quem retorna esses posts mais relevantes é a própria API do Twitter que tem suas próprias métricas.

7:13pm - Participante CL

Certo, agora ficou mais claro

7:13pm - Gustavo Fernandes

OK. Seguindo então...

Teria alguma outra sugestão de como esses dados poderiam ser coletados e trabalhados para que os resultados fossem aprimorados?

7:16pm - Participante CL

Só achei que essa relação não ficou clara de primeira, mas como você mesmo disse é a própria api do twitter que faz isso né? Então não sei.

Talvez uma explicação mais detalhada na página...

7:16pm - Gustavo Fernandes

Consegue imaginar alguma outra informação que poderia ser considerada e explorada na montagem dos perfis dos usuários?

7:18pm - Participante CL

Não imagino não

7:18pm - Gustavo Fernandes

O que você achou da interface do protótipo? Conseguiu trocar os perfis, selecionar intervalos de tempo e categorias, ver posts relacionados aos termos e assuntos de forma simples?

7:20pm - Participante CL

Consegui sim, tá tudo bem tranquilo e fácil de usar... na primeira vez que entrei tentei clicar nos títulos em vermelho de Texto, Link, etc achando que poderia ir direto pra lá... Então minha sugestão poderia ser de colocar um link ali talvez

7:21pm - Gustavo Fernandes

Entendi. Anotado.

Dentre os tipos de perfis apresentados, você acredita que algum conseguiu mapear de forma satisfatória o seu perfil?

7:22pm - Participante CL

Texto e Link sim, bastante... Já Trends ficou confuso por essa relação que o twitter faz, então não fica claro de primeira, só entrando lá na explicação e vendo o porquê da relação

7:23pm - Gustavo Fernandes

Conseguiria ordenar estes perfis do mais representativo para o menos representativo?

7:25pm - Participante CL

Link, Texto, Trend

7:25pm - Gustavo Fernandes

Os interesses trazidos na pesquisa do seu perfil mais representativo, no caso o LINK, realmente condizem com os seus interesses?

7:26pm - Participante CL
Eu diria que 80% sim

7:26pm - Gustavo Fernandes
Sentiu falta de algum?

7:28pm - Participante CL
Na verdade não, mas o 80% eu dei por ter coisas que eu na verdade retwitto e não considero exatamente "meu perfil"

7:29pm - Gustavo Fernandes
Entendi. É que o protótipo considera também os retweets porque de alguma forma acredita que você se interessa por aquele assunto.
Algum interesse encontrado que você considere ter sido inferido corretamente te surpreendeu?

7:31pm - Participante CL
Não

7:31pm - Gustavo Fernandes
O que você achou dos assuntos que foram inferidos no perfil que você considerou mais representativo?
Lembrando que os assuntos eram exibidos no gráfico de pizza.

7:34pm - Participante CL
Eu achei que conduzem bem com o meu perfil, talvez com um percentual não tão preciso mas bem próximo. Um exemplo é o Esporte, que correspondeu a 80% por eu postar os treinos de corrida que faço usando apps no celular...
Mas considerando que o universo analisado é apenas o twitter, então é isso mesmo

7:35pm - Gustavo Fernandes
Então você considera que são realmente assuntos que você possui interesse?

7:36pm - Participante CL
sim, com certeza

7:36pm - Gustavo Fernandes
Sentiu falta de algum assunto?

7:36pm - Participante CL
Não

7:37pm - Gustavo Fernandes
Você considera assertivo o mapeamento dos posts que deram origem aos seus interesses?

7:39pm - Participante CL
Na maioria sim, com ressalva apenas pra pessoa, que me colocou em primeiro lugar por eu postar fotos no instagram... mas aí acredito ser um detalhe que não tem muito que fazer

7:39pm - Gustavo Fernandes

Achou algum(ns) post(s) que não fazia(m) sentido para um determinado interesse ou assunto?

7:40pm - Participante CL

Não

7:40pm - Gustavo Fernandes

Como você vê a utilização desta abordagem para montar os perfis em outras redes sociais?

7:42pm - Participante CL

Acredito que possa ser até mais preciso. O twitter é um ótimo estudo, mas limitado pelo espaço. Em outras redes talvez possa conseguir muito mais informações ou no mínimo um universo maior para estudar

7:43pm - Gustavo Fernandes

Poderia citar alguns exemplos de outras redes sociais onde essa abordagem também poderia ser aplicada?

7:45pm - Participante CL

Eu acho que o LinkedIn e o facebook seria bons estudos.

7:45pm - Gustavo Fernandes

OK.

Acredita que em uma rede social a qual não possua limitação de caracteres para postar informações, os perfis inferidos serão mais representativos?

7:46pm - Participante CL

Com certeza

7:46pm - Gustavo Fernandes

Conseguiria explicar o por quê?

7:49pm - Participante CL

Por não limitar o conteúdo, a pessoa pode se expressar com mais liberdade sem ter que cortar palavras e escolher as menores. No twitter essa limitação acaba levando a pessoa a abreviações e frases cortadas que podem levar a uma margem de erro maior

7:49pm - Gustavo Fernandes

Entendi.

Encerramos por aqui. Muito agradecido pelo seu tempo e paciência.

7. Participante JG entrevistado no dia 09/04/2013 via Facebook Chat

5:57pm - Gustavo Fernandes

Você acessa a internet todos os dias?

5:58pm - Participante JG
Sim, todos os dias.

5:58pm - Gustavo Fernandes
Qual o dispositivo mais utilizado para acessar a internet?

5:59pm - Participante JG
iPhone e iPad

5:59pm - Gustavo Fernandes
Com que frequência faz uso de redes sociais?

5:59pm - Participante JG
Uso raramente (1x por mês) o Facebook e mais o Tweeter (2x por semana provavelmente)

6:00pm - Gustavo Fernandes
Você conseguiu compreender como a abordagem de montagem do perfil foi desenvolvida?

6:06pm - Participante JG
Sim. Você criou um mecanismo que sumariza por grandes grupos temáticos elementos (textos, links e trend topics) que são publicados no tweeter.

6:06pm - Gustavo Fernandes
Isso.
Teria alguma outra sugestão de como esses dados poderiam ser coletados e trabalhados para que os resultados fossem aprimorados?

6:09pm - Participante JG
Olhe, estive pensando que um campo de busca poderia ser legal para as opções de texto e trend topics
além disso, seria legal, em termos de sofisticação apenas, não de usabilidade, a adição de um tag cloud para texto e trend topics.

6:10pm - Gustavo Fernandes
Você diz um campo de busca para buscar por algum termo que possa ter sido trazido no seu perfil?
Sobre a tag cloud, entendido e anotado.

6:11pm - Participante JG
Exatamente. Seria um modo de rastrear livremente a base de textos que eu criei

6:12pm - Gustavo Fernandes
Entendido. Consegue imaginar alguma outra informação que poderia ser considerada e explorada na montagem dos perfis dos usuários?

6:12pm - Participante JG

Talvez, a possibilidade de realizar a busca por outros usuários possa também ser interessante. Porém, já onera muito o eventual engine de busca e poderia não ter tanto propósito dentro do objetivo da ferramenta

6:13pm - Gustavo Fernandes

OK. O que você achou da interface do protótipo? Conseguiu trocar os perfis, selecionar intervalos de tempo e categorias, ver posts relacionados aos termos e assuntos de forma simples?

6:14pm - Participante JG

Sim. Outra coisa que pensei é que poderiam existir alguns histogramas medindo publicação de tweets, acessos à pagina principal do Tweeter, retweets realizados em cima dos meus tweets (isso seria bem legal), comparação de novos tweets que eu crio versus os que eu retweeto.

"acessos à pagina principal do Tweeter" = meus acessos

6:15pm - Gustavo Fernandes

entendido.

Tem alguma sugestão de recursos que poderiam ser adicionados à interface de consulta fora os que você mencionou acima?

6:17pm - Participante JG

Sobre a interface/usabilidade, gostei muito. Bem descomplicada e direto ao ponto. Chamou-me a atenção a preocupação com a padronagem das fontes (cor, tamanho, tipo). Achei a padronagem usada muito apropriada e bem pensada

6:18pm - Gustavo Fernandes

Dentre os tipos de perfis apresentados, você acredita que algum conseguiu mapear de forma satisfatória o seu perfil?

Lembrando que os tipos são: TEXTO, LINK e TRENDING TOPIC.

6:20pm - Participante JG

Uma sofisticação, imagino que bastante avançada, seria a possibilidade de saber o que outros usuários estão tweetando com respeito a temas que eu já tweetei, por exemplo, usuário notórios e publicos, e haver uma correlação entre o que estou publicando e retweetando e o que um determinado grupo notório (políticos, artistas, religiosos, filósofos, jornalistas, escritores, etc.) também está. Esse poderia ser um indicador de o quanto o usuário está "atenado" com assuntos em evidência ou dominantes no status quo.

6:21pm - Gustavo Fernandes

Entendi. Realmente seria interessante.

Mas voltando a pergunta: dentre os tipos de perfis apresentados, você acredita que algum conseguiu mapear de forma satisfatória o seu perfil?

Lembrando que os tipos são: TEXTO, LINK e TRENDING TOPIC.

6:24pm - Participante JG

Sobre o mapeamento, sim, o gráfico do de texto reflete bem o tipo de assuntos que costumo consumir/produzir no tweeter. Eu faria apenas uma ressalva para a parte de esportes, pois eu sigo muito e gero muita coisa sobre MMA, UFC p.ex.), e,

acredito, a fatia dos esportes deveria ser maior por isso. Mas isso é apenas um feeling.

6:24pm - Gustavo Fernandes

Entendi.

Mas quais tipos de perfis conseguiram te representar corretamente? Todos?

6:27pm - Participante JG

O de texto achei o mais acurado, em primeiro lugar. Em segundo o de links.

6:28pm - Gustavo Fernandes

Então você considera que o TEXTO e o LINK foram os que conseguiram te representar de forma correta, certo?

6:28pm - Participante JG

sim. O TT não me pareceu muito condizente

6:29pm - Gustavo Fernandes

Os interesses trazidos na pesquisa do seu perfil mais representativo realmente condizem com os seus interesses?

Lembrando que os interesses estavam listados na tabela a direita.

6:31pm - Participante JG

Sim, condizem. Mas levo em conta para isso os labels usados no gráfico, e não os itens da tabela, que são, "taxonomicamente", diferentes da classificação usada no gráfico.

6:31pm - Gustavo Fernandes

OK. Sentiu falta de algum interesse?

6:34pm - Participante JG

Sim.

Acho que poderiam haver duas quebras adicionais:

1 - Esportes: talvez fosse bacana haver uma quebra por futebol (que deve ser dominante) e os demais, e haver ainda alguma possibilidade de drill down (isso já remete às sugestões de usabilidade que falei acima)

2 - Notícias internacionais: acho que seria interessante o usuário conhecer o que ele consome/produz com respeito a temas de sua localidade (em termos nacionais) e fora dela.

6:36pm - Gustavo Fernandes

Entendido.

Algum interesse encontrado que você considere ter sido inferido corretamente te surpreendeu?

6:37pm - Participante JG

Achei interessante o "temas policiais". Porém, os tweets dentro deste grupo não são desta temática

6:38pm - Gustavo Fernandes

E sobre os termos da tabela da direita, algum te surpreendeu?

6:39pm - Participante JG
(preciso de 2 min p/ continuar)

6:39pm - Gustavo Fernandes
OK.
Te aguardo.

6:51pm - Participante JG
mais um min

6:51pm - Gustavo Fernandes
tranquilo.

6:56pm - Participante JG
oi gustavo. voltei

6:56pm - Gustavo Fernandes
OK.
voltando...
Sobre os termos listados na tabela de interesses, algum te surpreendeu?

6:59pm - Participante JG
nao. achei que foram bem condizentes com o que escrevo

6:59pm - Gustavo Fernandes
Você chegou a comentar sobre a fatia de esportes que deveria ser maior nos assuntos, mas de forma geral o que você achou dos assuntos que foram inferidos no perfil que você considerou mais representativo?
Os assuntos são exibidos no gráfico.

7:04pm - Participante JG
achei a inferência muito boa no sentido de determinação dos grupos de temas. Porém, ela ainda deve ser aprimorada para não colocar dentro dos grupos tweets que sejam flagrantemente não associados àquele grupo. P.ex., no grupo ciência e tecnologia há o seguinte tweet: "@SalimMaluf Dr. Paulo, atenção ao Mollusco, pois daqui a pouco ele comparará o sr. ao Ali Babá ou, ainda pior, ao Zé Dirceu! aceite isso!!". Este aqui deveria estar em política.

7:05pm - Gustavo Fernandes
OK.
Sentiu falta de algum assunto?

7:08pm - Participante JG
Importante: sobre o meu comentário acima, eu estava me referindo aos resultados no perfil de texto. Já no perfil link, as associações tweet X tema foram muito, muito mais precisas
não. Os assuntos estão muito bem mapeados.

7:09pm - Gustavo Fernandes

Você considera assertivo o mapeamento dos posts que deram origem aos seus interesses?

7:11pm - Participante JG

Eu nao entendi. O que seriam os posts?

7:11pm - Gustavo Fernandes

posts são as mensagens que você inseriu quando você "posta" uma mensagem no twitter, estou chamando de post. um post = uma mensagem inserida no Twitter

E a pergunta é se você considera assertivo o mapeamento dos posts que deram origem aos seus interesses?

Você consegue saber isso quando clica em um termo e abre um popup listando todos os posts que deram origem àquele termo.

O mesmo acontece quando clica-se em uma das categorias do gráfico.

7:17pm - Participante JG

Certo.

Considero pouco/médio assertivo esse mapeamento no perfil texto.

Considero bem assertivo na parte de links

7:18pm - Gustavo Fernandes

Achou algum(ns) post(s) que não fazia(m) sentido para um determinado interesse ou assunto?

7:21pm - Participante JG

Sim, eu até mencionei um exemplo acima. Nesse comentário:

7:21pm - Participante JG

achei a inferência muito boa no sentido de determinação dos grupos de temas. Porém, ela ainda deve ser aprimorada para não colocar dentro dos grupos tweets que sejam flagrantemente não associados àquele grupo. P.ex., no grupo ciência e tecnologia há o seguinte tweet: "@SalimMaluf Dr. Paulo, atenção ao Mollusco, pois daqui a pouco ele comparará o sr. ao Ali Babá ou, ainda pior, ao Zé Dirceu! aceite isso!!". Este aqui deveria estar em política.

7:22pm - Gustavo Fernandes

é verdade.

tá quase no fim

Como você vê a utilização desta abordagem para montar os perfis em outras redes sociais?

7:22pm - Participante JG

blz

7:22pm - Gustavo Fernandes

Como você vê a utilização desta abordagem para montar os perfis em outras redes sociais?

7:24pm - Participante JG

Eu acho muito bacana. Acho que ela pode ser muito útil para mapeamento de tendências e comportamento de usuários com respeito a entidades, pessoas, empresas, marcas, etc.

7:24pm - Gustavo Fernandes

Você acredita que a abordagem utilizada para a montagem dos perfis também seria extensível para outras redes sociais?

7:27pm - Participante JG

Acho que sim. A classificação por trending topics, evidentemente não. Mas link e texto com certeza.

No caso do facebook, seria util haver também uma classificação de videos publicados

7:27pm - Gustavo Fernandes

Entendi.

Acredita que em uma rede social a qual não possua limitação de caracteres para postar informações, os perfis inferidos serão mais representativos?

7:29pm - Participante JG

Acredito que sim. Acho quase certo que a maior variedade de vocábulos e a menor necessidade de abreviações tornarão o mecanismo de inferência muito mais preciso

7:30pm - Gustavo Fernandes

Chegamos ao fim. Muito agradecido pelo seu tempo e paciência.

8. Participante ET entrevistado no dia 13/04/2013 via Facebook Chat

10:19am - Gustavo Fernandes

Você acessa a internet todos os dias?

10:19am - Participante ET

sim

10:19am - Gustavo Fernandes

Qual o dispositivo mais utilizado para acessar a internet?

10:20am - Participante ET

hoje diria que o acesso via celular esta quase igual ao acesso via laptop, mas o laptop ainda ganha por pouco

PAUSA

precisa ser respostas SIM ou NAO ou posso falar mais?

10:21am - Gustavo Fernandes

o objetivo é você falar livremente

Com que frequência faz uso de redes sociais?

10:21am - Participante ET
Muita
diariamente
horariament
minutamente

10:21am - Gustavo Fernandes
hehehe
Quais?

10:21am - Participante ET
facbeook, twitter
linkedin

10:22am - Gustavo Fernandes
OK.
Vamos falar agora um pouco sobre as informações que foram inferidas na montagem do seu perfil.

10:22am - Participante ET
ok

10:23am - Gustavo Fernandes
Você conseguiu compreender como a abordagem de montagem do perfil foi desenvolvida?

10:23am - Participante ET
sim
porem
achei que a classificacao dos temas estava confusa

10:24am - Gustavo Fernandes
entendi.
Teria alguma outra sugestão de como esses dados poderiam ser coletados e trabalhados para que os resultados fossem aprimorados?

10:24am - Participante ET
uma classificacao mais refinada de temas
cito um exemplo
tinham 2 categorias pra mim muito similares
esporte e jogos
(no meu entendimento)
em uma delas, tinham tweets meus sobre neto baiano, que na verdade podem ser classificados nos 2
haaaa

10:25am - Gustavo Fernandes
entendi. Na verdade sua crítica é em relação aos assuntos que são mostrados no gráfico, correto?

10:25am - Participante ET
pq uso o termo artilheiro do Brasil
!
ai
ela foi classificada nao como esporte
e sim como Brasil
certo
portanto acho que pra classificar os tweets talvez precise analisar mais termos e descobrir correlacoes entre eles

10:26am - Gustavo Fernandes
entendi. anotado.
Mas e em relação aos interesses que são mostrados na tabela da direita?

10:27am - Participante ET
ai sim,
me vi ali
claramente vi temas que efetivamente tweeto

10:27am - Gustavo Fernandes
OK. Consegue imaginar alguma outra informação que poderia ser considerada e explorada na montagem dos perfis dos usuários?

10:27am - Participante ET
nao

10:28am - Gustavo Fernandes
O que você achou da interface do protótipo? Conseguiu trocar os perfis, selecionar intervalos de tempo e categorias, ver posts relacionados aos termos e assuntos de forma simples?

10:28am - Participante ET
sim, boa usabilidade. fez com a Huga?

10:28am - Gustavo Fernandes
hahaha... nem chega perto!
Tem alguma sugestão de recursos que poderiam ser adicionados à interface de consulta?

10:29am - Participante ET
volto ao meu ponto acima: talvez pudesse ter uma forma de ver a combinacao dos termos
pois me vi, como no exemplo citado, com varios tweets que mistruam assuntos
o exemplo la do neto baiano artilheiro do brasil, ficou so em brasil e na verdade eh mais espote(mesmo tb podendo ser considerado brasil)

10:31am - Gustavo Fernandes
Beleza.
Dentre os tipos de perfis apresentados, você acredita que algum conseguiu mapear de forma satisfatória o seu perfil?

Lembrando que os tipos eram TEXTO, LINK e TRENDING TOPIC

10:31am - Participante ET

o primeiro mais, mas o segundo tb

nao uso tanot TT, por isso nao me achei tanto

texto foi o melhor, com as restricoes de classificacoes que ja falei

10:32am - Gustavo Fernandes

Entendi. Então pelo que você falou, a ordem de representatividade do perfil seria Texto, Link e TT?

10:32am - Participante ET

certo

10:33am - Gustavo Fernandes

TT você já falou que não conseguiu montar o seu perfil pela escassez de trending topics, mas texto e link, você acredita que representou bem o seu perfil?

10:34am - Participante ET

sim

talvez tenha sido impressao

mas me vi mais nos 2 primeiros

10:35am - Gustavo Fernandes

Iria te perguntar sobre se os interesses listados representavam bem o seu perfil, mas você já confirmou que sim.

10:35am - Participante ET

isso

sem duvida

10:35am - Gustavo Fernandes

Sentiu falta de algum interesse?

10:35am - Participante ET

nao

me surpreendi com alguns

mas ai talvez seja a questao da classificao

ue por uma palavra so eh dificil de classificar preciso

10:36am - Gustavo Fernandes

Essa era a minha próxima pergunta: algum interesse encontrado que você considere ter sido inferido corretamente te surpreendeu?

10:36am - Participante ET

eu penso na frente

sim

tipo Organizacoes

e tinha uma palavra la chamada Rafael Gomes (em numero 10 da lista) que eh um dev da huge que so citei 1 vez na vida

10:37am - Gustavo Fernandes

hehehe

OK.

Sobre os assuntos você já fez algumas observações, mas de maneira geral os assuntos que foram inferidos no perfil que você considerou mais representativo (TEXTO) são assuntos que você se interessa?

10:39am - Participante ET

sim

com as consideracoes que ja coloquei

10:39am - Gustavo Fernandes

OK. Sentiu falta de algum assunto?

10:39am - Participante ET

no

10:39am - Gustavo Fernandes

Você considera assertivo o mapeamento dos posts que deram origem aos seus interesses?

10:39am - Participante ET

sim

10:40am - Gustavo Fernandes

Achou algum(ns) post(s) que não fazia(m) sentido para um determinado interesse ou assunto?

10:40am - Participante ET

sim, o de neto baiano

que descrevi

10:40am - Gustavo Fernandes

é verdade.

10:40am - Participante ET

e o fato do rafael gomes q so falei uma vez pra zoar estar entre as 10

10:40am - Gustavo Fernandes

entendi.

Como você vê a utilização desta abordagem para montar os perfis em outras redes sociais?

10:41am - Participante ET

ajuda a predizer meus gostos

10:41am - Gustavo Fernandes

Você acredita que a abordagem utilizada para a montagem dos perfis também seria extensível para outras redes sociais?

10:41am - Participante ET
que gosto de futebol e politica
entao de uma certa forma, isso ajuda a me encaixar em grupos de perfis similares
foi?

10:43am - Gustavo Fernandes
Entendi, mas em relação à aplicabilidade desta abordagem para a montagem dos
perfis em outras redes sociais?

10:43am - Participante ET
as crianças estão se degladiando aqui

10:43am - Gustavo Fernandes
faltam só duas

10:43am - Participante ET
acho aplicavel, e extremamente util

10:43am - Gustavo Fernandes
Poderia citar alguns exemplos?

10:43am - Participante ET
blz
sim:
iamgina uma rede social como o pinterest
se com essa ferramenta sua ja consegue prever o que eh meu interesse, isso pode
automaticamente configurar o pinterest pra me mostrar o que quero/ gosto de ver
e falar

10:44am - Gustavo Fernandes
entendi.
Acredita que em uma rede social a qual não possua limitação de caracteres para
postar informações, os perfis inferidos serão mais representativos?

10:45am - Participante ET
poderia ser bem diferente pelo motivo que citei: se a classificacao dos tweets nao
leva em consideracao uma combinacao de palavras, quanto mais palavras (falta de
limite de caracteres) maior a chance de ser impreciso na classificao
vide exemplo neto baiano
se eu pudesse falar mais dele, sei la talvez da cidade onde nasceu, a classificao
poderia ser outra e totlamente imprecisa

10:46am - Gustavo Fernandes
entendi.
É isso. Encerramos aqui. Muito agradecido pelo seu tempo e paciência!

9. Participante RG entrevistado no dia 13/04/2013 via Facebook Chat

1:18pm - Gustavo Fernandes
Você acessa a internet todos os dias?

1:19pm - Participante RG
Sim. Todos os dias. E passo aproximadamente 18 horas conectado.

1:19pm - Gustavo Fernandes
Qual o dispositivo mais utilizado para acessar a internet?

1:21pm - Participante RG
Smartphone.
Mas também acesso através do Notebook profissional.

1:21pm - Gustavo Fernandes
Com que frequência faz uso de redes sociais?

1:22pm - Participante RG
todo dia. às vezes consulto o Facebook até 5 vezes.
(hoje já é a 3a... neste ritmo...rs)

1:22pm - Gustavo Fernandes
Só o Facebook e o Twitter ou mais alguma?

1:23pm - Participante RG
Facebook, Twitter, Google+ e Instagram.

1:23pm - Gustavo Fernandes
OK.
Vamos falar agora um pouco sobre as informações que foram inferidas na montagem do seu perfil.
Você conseguiu compreender como a abordagem de montagem do perfil foi desenvolvida?

1:24pm - Participante RG
Acho que entendi sim.
Mas tenho alguns observações
Já posso falar?

1:25pm - Gustavo Fernandes
acho que durante as perguntas essas observações vão se encaixando...
Teria alguma outra sugestão de como esses dados poderiam ser coletados e trabalhados para que os resultados fossem aprimorados?

1:27pm - Participante RG
Atentar para que não seja colhidas palavras que não estão relacionadas com o próprio Tweet. As vezes percebi q algumas palavras estavam dentro do link que twitava, mas sem nada a ver com o meu objetivo
Era palavras no HTML dos links divulgados
Talvez o algoritmo devêssemos ater nos títulos das páginas chamadas

1:29pm - Gustavo Fernandes

entendi, mas o objetivo do protótipo era justamente este, ou seja, montar três perfis distintos, um considerando somente o texto, outro o texto presentes nos links e outros os textos dos trending topics

Vamos chegar a parte das perguntas sobre a sua questão da avaliação de cada tipo de perfil.

Consegue imaginar alguma outra informação que poderia ser considerada e explorada na montagem dos perfis dos usuários?

1:31pm - Participante RG

No momento não

1:31pm - Gustavo Fernandes

OK. O que você achou da interface do protótipo? Conseguiu trocar os perfis, selecionar intervalos de tempo e categorias, ver posts relacionados aos termos e assuntos de forma simples?

1:32pm - Participante RG

Consegui, muito legal

1:33pm - Gustavo Fernandes

Tem alguma sugestão de recursos que poderiam ser adicionados à interface de consulta?

1:33pm - Participante RG

No momento nada me ocorre, mas posso depois, se tiver, te passar por email

1:34pm - Gustavo Fernandes

Não, sem problemas. Não precisa passar por e-mail.

Dentre os tipos de perfis apresentados, você acredita que algum conseguiu mapear de forma satisfatória o seu perfil?

Os tipos eram: TEXTO, LINK e TRENDING TOPIC.

1:35pm - Participante RG

TEXTO

1:36pm - Gustavo Fernandes

Conseguiria ordenar estes perfis do mais representativo para o menos representativo?

1:37pm - Participante RG

Sim

1:38pm - Gustavo Fernandes

Então seria em primeiro lugar TEXTO e depois?

1:39pm - Participante RG

Links

1:39pm - Gustavo Fernandes

E por último TT.

OK.

Os interesses trazidos na pesquisa do seu perfil mais representativo, o TEXTO, realmente condizem com os seus interesses?

Lembrando que os interesses estão listado na tabela a direita.

1:41pm - Participante RG

Acho que condizem com os interesses baseados nos Tweets, infelizmente neste momento não tenho como abrir para te dar certeza, falo pelo que lembro mais ou menos o q vi

1:42pm - Gustavo Fernandes

Entendi. Mas realmente o protótipo trata somente as informações que ele buscou no seu perfil do Twitter.

Sentiu falta de algum (interesse)?

1:43pm - Participante RG

Baseado no Twitter acredito q não.

1:43pm - Gustavo Fernandes

Algum interesse encontrado que você considere ter sido inferido corretamente te surpreendeu?

1:43pm - Participante RG

Sim

1:43pm - Gustavo Fernandes

Qual(is)?

1:44pm - Participante RG

Renato Calheiros... Rs

1:44pm - Gustavo Fernandes

hehehe

O que você achou dos assuntos que foram inferidos no perfil que você considerou mais representativo?

Lembrando que os assuntos estão listados no gráfico de pizza.

1:45pm - Participante RG

Meu interesse na é é ve-lo fora do senado

Os relacionados a tecnologia que apareceram

1:46pm - Gustavo Fernandes

interessante essa tua observação porque uma das propostas de trabalho suturo é fazer análise de sentimento nestes termos, ou seja, não é porque apareceu algo que você se interesse que isso você considera positivo, afinal coisas que não gostamos também são mencionadas.

suturo não, futuro

1:47pm - Participante RG

Muito bom

1:48pm - Gustavo Fernandes

Mas sobre os assuntos, você considerou que os assuntos que foram inferidos no seu perfil são realmente assuntos de seu interesse?

Assuntos estão listados no gráfico de pizza.

1:54pm - Participante RG

Alguns sim

Cara, vou ter q dirigir, falta muito?

1:55pm - Gustavo Fernandes

umas 4 perguntas.

se quiser terminamos depois.

1:56pm - Participante RG

Melhor, pois assim abro meu perfil e te respondo vendo, Ok?

1:57pm - Gustavo Fernandes

beleza.

Vou ficar online no FB e depois quando puder e se ainda estiver online, me chama que terminamos.

1:58pm - Participante RG

Ok!

1:58pm - Gustavo Fernandes

valeu

3:38pm - Participante RG

tá por ae?

3:57pm - Gustavo Fernandes

voltei

Vamos finalizar?

3:58pm - Participante RG

bora.

passa novamente por favor o link do meu perfil.

não precisa, achei

3:58pm - Gustavo Fernandes

<http://sal.uniriotec.br/dpm/pre-home?user=xxxxxxxx>

3:59pm - Participante RG

manda

3:59pm - Gustavo Fernandes

Estávamos falando sobre os assunto.

assuntos

O que você achou dos assuntos que foram inferidos no perfil que você considerou mais representativo?

Que foi o perfil TEXTO.

4:00pm - Participante RG
me explica antes uma coisa
esta ultima combo-box, como vc a alimenta?
é pre-definida?

4:01pm - Gustavo Fernandes
São tipos ou categorias já pré-definidas.
isso
Tudo que não se encaixa nas 4 primeiras categorias entra na categoria OUTROS.

4:03pm - Participante RG
pera to dando uma olhada

4:03pm - Gustavo Fernandes
OK

4:03pm - Participante RG
cara
seu projeto é muito interessante.
realmente.
O fato é que utilizei realmente o twitter para divulgar links de interesse e a maioria está relacionado realmente à Artes e Entretenimento, um dos meus maiores interesses
no perfil Texto acusou de fato isto. Com 47%
Em segundo lugar, gosto muito de temas filosóficos, religisos, história das religiões

4:06pm - Gustavo Fernandes
Entendi. Sentiu falta de algum assunto?

4:06pm - Participante RG
e mais uma vez o DPM fez um gol onde a coruja dorme. em 2o lugar com quase 17%
Para vc ver: eu tenho dois portais na internet relacionados à artes. Tenho um grupo aqui no facebook sobre filmes.

4:07pm - Gustavo Fernandes
Legal.
Você considera assertivo o mapeamento dos posts que deram origem aos seus interesses?

4:08pm - Participante RG
destes, acho que realmente são os assuntos que mais me interessa, isto se vc me disser que música está dentro de Artes e Entretenimento, não existe este tema à parte.
confirmado isto, está OK

Senti falta de um assunto, mas acho que não twitei muito mesmo sobre a indústria do casamento, devido ao outro portal

4:09pm - Gustavo Fernandes

ah tá. Realmente essa categoria não existe.

Você considera assertivo o mapeamento dos posts que deram origem aos seus interesses?

4:10pm - Participante RG

acho que sim

4:10pm - Gustavo Fernandes

Achou algum(ns) post(s) que não fazia(m) sentido para um determinado interesse ou assunto?

E se achou conseguiria citar algum?

4:11pm - Participante RG

me dê poucos minutos

4:11pm - Gustavo Fernandes

Fique a vontade.

4:15pm - Participante RG

Sugestão: na lista de interesses poderia aparecer o número de vezes que cada termo apareceu

4:15pm - Gustavo Fernandes

OK. Anotado.

Além do número de vezes também existe uma relevância que é um número percentual.

4:17pm - Participante RG

Cara, não tem mesmo nenhum termo que aparece mais que estes...? selecionei tudo do perfil Texto

é que falo muito de Instagram também...

4:18pm - Gustavo Fernandes

Entendi. De repente apareceu, mas o algoritmo não deve ter julgado como uma entidade relevante.

Seguindo...

Como você vê a utilização desta abordagem para montar os perfis em outras redes sociais?

4:19pm - Participante RG

Acho que é um caminho razoável. E acredito que algo similar já esteja fazendo isto no G+ e FB

Talvez no Twitter

o que me chamou mais atenção no seu, foi o fato do usuário usar

4:20pm - Gustavo Fernandes

É bem possível que sim, mas é algo intrínseco da plataforma deles e não expõe esses dados.

Acredita que em uma rede social a qual não possua limitação de caracteres para postar informações, os perfis inferidos serão mais representativos?

4:21pm - Participante RG

exato. como futuro empresário, estou imaginando o que este tipo de análise para o próprio usuário poderia ser aproveitado comercialmente

4:21pm - Gustavo Fernandes

hehehe

4:21pm - Participante RG

Sim, acredito.

4:21pm - Gustavo Fernandes

Conseguiria explicar por que?

Por que seria mais representativo?

4:22pm - Participante RG

com mais conteúdo para o algoritmo trabalhar, melhor. O pulo do gato é fazer um algoritmo com regras essenciais para a classificação do que é mais relevante para cada usuário

4:23pm - Gustavo Fernandes

Entendi.

Chegamos ao fim. Muito agradecido pelo seu tempo e paciência.

4:23pm - Participante RG

que isto, foi um prazer, ainda mais para este trabalho. Parabéns. Agora uma pergunta para você

4:23pm - Gustavo Fernandes

fale

4:24pm - Participante RG

já está pensando em colocar isto em prática fora do mundo acadêmico?

Estou trabalhando em 8 projetos que serão apresentados para alguns investidores talvez eu consiga encaixar este também. cada projeto antes era uma startup, agora é como se fosse uma com portais onde existe sinergia imaginei este algoritmo trabalhando em cima dos perfis dos membros de cada portal

10. Participante WS entrevistado no dia 16/04/2013 via Google Talk

2:55 PM

Gustavo Fernandes: Você acessa a internet todos os dias?

Participante WS: Sim.

Gustavo Fernandes: Qual o dispositivo mais utilizado para acessar a internet?
Participante WS: Desktop.

2:56 PM

Gustavo Fernandes: Com que frequência faz uso de redes sociais?
Participante WS: Diariamente.

Gustavo Fernandes: Quais?
Participante WS: Twitter.
Só o twitter.

Gustavo Fernandes: Vamos falar agora um pouco sobre as informações que foram inferidas na montagem do seu perfil.
Você conseguiu compreender como a abordagem de montagem do perfil foi desenvolvida?

2:57 PM

Participante WS: Acho que sim.

Gustavo Fernandes: Teria alguma outra sugestão de como esses dados poderiam ser coletados e trabalhados para que os resultados fossem aprimorados?

2:58 PM

Participante WS: Qual seria o objetivo para mim, como usuário?

2:59 PM

Gustavo Fernandes: Na verdade o objetivo do protótipo foi tentar demonstrar os seus interesses através do processamento automático das mensagens que você postou no twitter em 2012.

Delas que esses termos e assuntos foram extraídos.

Participante WS: Quer dizer, assim eu não saberia como sugerir alguma coisa, sem saber como eu poderia usar. Me parece que a ferramenta tem como alvo alguém fazendo alguma pesquisa em cima dos dados do usuário.

Ok. Deixa eu ver aqui então. 1 minuto.

Não me ocorre nada por agora.

3:01 PM

Gustavo Fernandes: seguindo então...

Consegue imaginar alguma outra informação que poderia ser considerada e explorada na montagem dos perfis dos usuários?

3:05 PM

Participante WS: A quantidade de twitts usados para traçar o perfil.
Com esse dado vc pode fazer alguma coisa relacionada a frequência.
Não sei se seria útil...

Gustavo Fernandes: entendi.

O que você achou da interface do protótipo? Conseguiu trocar os perfis, selecionar intervalos de tempo e categorias, ver posts relacionados aos termos e assuntos de forma simples?

3:08 PM

Participante WS: Sim. A interface está simples e direta. Não tive nenhuma dificuldade com ela.

Gustavo Fernandes: Teria alguma sugestão de recursos que poderiam ser adicionados à interface de consulta?

3:09 PM

Participante WS: Um filtro por datas (períodos) pode ser interessante.

3:10 PM

Gustavo Fernandes: Entendi. Em vez de separar por quartos, correto? aliás por quadrimestres.

3:11 PM

Participante WS: Sim. Os quartos são diretos e servem como um "preset". Mas poder selecionar um determinado período, como um mês em particular de um ano qualquer pode dar mais flexibilidade a montagem dos perfis. Ou não...

3:12 PM

Gustavo Fernandes: Entendido e anotado.

Participante WS: Ou um período qualquer entre duas datas...

Gustavo Fernandes: Beleza.

Dentre os tipos de perfis apresentados (TEXTO, LINK ou TRENDING TOPIC), você acredita que algum conseguiu mapear de forma satisfatória o seu perfil?

3:14 PM

Participante WS: Realmente o filtro pega todos os principais assuntos que andei falando, mas dependendo do caso não necessariamente expressa meu interesse.

Por exemplo:

No perfil TEXTO aparece 15,2% de "Religião". Normalmente eu não tenho interesse nenhum nisso, mas me chama a atenção a quantidade de absurdos ligados a esse assunto, o que acaba provocando algum comentário da minha parte. No final das contas se eu comentei é pq houve algum interesse, negativo ou positivo.

Então, respondendo a sua pergunta: sim.

Tá aí ainda?

3:19 PM

Gustavo Fernandes: to aqui entendi

na verdade você considera então que o perfil TEXTO foi o que melhor te representou, correto?

Participante WS: Sim.

Gustavo Fernandes: Algum outro você considere que tenha te representando corretamente?

LINK ou TRENDING TOPIC?

3:22 PM

Participante WS: LINK sim. TRENDING TOPIC, eu tenho quase nenhuma atividade, o pouco que tem não me pareceu correto.

Gustavo Fernandes: Então pela ordem ficou TEXTO, LINK e TRENDING TOPIC, correto?

Participante WS: Sim.

Acho que vc precisa dar um olhado no filtro dos trending topics.
uma olhada

Gustavo Fernandes: blz. Os outros usuários tb reportaram essa dificuldade de trending topic.

Os interesses trazidos na pesquisa do seu perfil mais representativo realmente condizem com os seus interesses?

Lembrando que os interesses são listados na tabela a direita do posição 1 a 10.
e no caso o seu perfil mais representativo foi o TEXTO.

3:25 PM

Participante WS: Sim. Acho que está bem classificado.

Gustavo Fernandes: Sentiu falta de algum interesse?

3:26 PM

Participante WS: Senti, mas tem mais haver com pessoas ou blogs ou apps que eu sigo do que com o que eu escrevo.

Então não sei se aplica.

3:27 PM

Gustavo Fernandes: Entendi. Mas realmente o perfil montado foi somente em cima do que você postou no Twitter.

3:28 PM

Participante WS: Certo.

Gustavo Fernandes: Algum interesse encontrado que você considere ter sido inferido corretamente te surpreendeu?

Participante WS: Política e negócios.

3:29 PM

Gustavo Fernandes: Legal.

O que você achou dos assuntos que foram inferidos no perfil que você considerou mais representativo?

Lembrando que os assuntos são exibidos no gráfico

3:30 PM

Participante WS: Eles variam de perfil para perfil, ou esta divisão é fixa?

Gustavo Fernandes: variam de perfil para perfil

na verdade variam a cada filtro que você aplica (por tempo e categoria, inclusive).

3:32 PM

Participante WS: Eles pegaram de forma abrangente todos os interesses. Não ficou muito claro pq "Cultura e Política" estão juntos.

Gustavo Fernandes: Entendi. Então como sugestão você gostaria de separar cultura e política.

Sentiu falta de algum assunto?

3:33 PM

Participante WS: Não. Tudo de relevante está lá.

3:34 PM

Gustavo Fernandes: Você considera assertivo o mapeamento dos posts (mensagens inseridas) que deram origem aos seus interesses?

É possível ver isso quando algum item do gráfico ou algum termo da tabela é clicado.

Participante WS: Sim. Eu estou só olhando aqui rapidinho.

3:36 PM

Gustavo Fernandes: OK. Sem problemas, te espero.

3:37 PM

Participante WS: Sim, se com assertivo vc quer dizer que o twiit se enquadrou na categoria.

Gustavo Fernandes: isso

ou no termo, no caso da tabela a direita

3:38 PM

Participante WS: O que eu vi nos dois, me pareceu correto.

Exceto nos trendings.

Gustavo Fernandes: OK.

Achou algum(ns) post(s) que não fazia(m) sentido para um determinado interesse ou assunto?

3:39 PM

Participante WS: Nos trendings.

Gustavo Fernandes: Conseguiria citar algum?

3:40 PM

Participante WS: Tem um termo nos Intereeses que é "Gov Kasich".

O twitt que gerou esse termo diz:

@danieldrehmer: Adeus devassa do gynshop <http://t.co/c8WC2JHP> // Carai!!!
#NotCool

Gustavo Fernandes: entendi.

Como você vê a utilização desta abordagem para montar os perfis em outras redes sociais?

Outras redes que não sejam o Twitter.

3:42 PM

Participante WS: Acho que pode funcionar muito bem, mas para redes como Facebook talvez vc tenha que expandir os perfis.

Eu não sou indicado para falar de outras redes, mas imagino que o twitter seja o que tem de mais "direto ao ponto", por causa da saudável limitação de 140 caracteres.

Gustavo Fernandes: A próxima pergunta é justamente sobre isso. Acredita que em uma rede social a qual não possua limitação de caracteres para postar informações, os perfis inferidos serão mais representativos?

3:45 PM

Participante WS: Acredito que sim, no entanto daria mais trabalho para filtrar.

3:46 PM

Gustavo Fernandes: Encerramos por aqui. Muito agradecido pelo seu tempo e paciência.