

# Universidade Federal do Rio de Janeiro

Bianca Munaro Lima

AIDCHAT: UMA ABORDAGEM DE RECOMENDAÇÃO DE CONTEÚDO EM SISTEMAS DE BATE-PAPO







# UNIVERSIDADE FEDERAL DO RIO DE JANEIRO INSTITUTO DE MATEMÁTICA INSTITUTO TÉRCIO PACITTI DE APLICAÇÕES E PESQUISAS COMPUTACIONAIS PROGRAMA DE PÓS-GRADUAÇÃO EM INFORMÁTICA

Bianca Munaro Lima

# AIDCHAT: UMA ABORDAGEM DE RECOMENDAÇÃO DE CONTEÚDO EM SISTEMAS DE BATE-PAPO

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Informática, Instituto de Matemática e Instituto Tércio Pacciti, Universidade Federal do Rio de Janeiro, como requisito parcial à obtenção do título de Mestre em Informática.

Orientador: Prof. Maria Luiza Machado Campos, Ph.D em Ciência da Computação

#### Bianca Munaro Lima

# AIDCHAT: UMA ABORDAGEM DE RECOMENDAÇÃO DE CONTEÚDO EM SISTEMAS DE BATE-PAPO

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Informática, Instituto de Matemática e Instituto Tércio Pacciti, Universidade Federal do Rio de Janeiro, como requisito parcial à obtenção do título de Mestre em Informática.

Aprovada em Rio de Janeiro,	21 de Janeiro de 2014.
	Prof <sup>a</sup> . Maria Luiza Machado Campos, Ph.D. – PPGI/UFRJ
	Prof <sup>a</sup> . Jonice de Oliveira Sampaio, Ph.D. – PPGI/UFRJ
	Prof <sup>a</sup> . Mariano Pimentel, D.Sc. – PPGI/UFRJ

# **Agradecimentos**

Agradeço aos meus pais, Lúcia e Jorge, que me ensinaram valores e sempre estiveram ao meu lado e ao meu irmão Felipe.

Ao meu noivo, Rogério, que me incentivou e me apoiou nos momentos mais difíceis e compreendeu meus momentos de ausência.

À minha orientadora, professora Maria Luiza Machado Campos, pela confiança em meu trabalho, pelo apoio e conselhos preciosos.

A todos os meus amigos, pelo incentivo e pela força e por entender os momentos em que não pude estar com eles.

A todas as pessoas que participaram dos experimentos com tanto dedicação e interesse, contribuindo para a realização desse trabalho.

Ao CAPES (Coordenação de Aperfeiçoamento de Pessoas de Nível Superior) pela concessão da bolsa durante um período de realização da pesquisa.

#### Resumo

LIMA, Bianca Munaro. 2013. 137 f. Dissertação (Mestrado em Informática) - Programa de Pós-Graduação em Informática, Instituto de Matemática, Instituto Tércio Pacitti, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2013.

A popularização da Web 2.0 provocou um considerável aumento na utilização dos meios de comunicação textual. Estes são de fácil uso e permitem discussões síncronas e assíncronas. Eles tornam possível a interação entre pessoas de qualquer parte do mundo rapidamente, revolucionando as formas de comunicação. No entanto, esse tipo de comunicação também pode apresentar dificuldades, principalmente quando envolve exclusivamente comunicação textual, pois os usuários utilizam apenas a troca de mensagens na forma de texto, não podendo se ver ou falar. Portanto, é comum que existam problemas como: ambiguidades na interpretação, dificuldades de entendimento, dificuldades de acompanhar a sequencia da conversa, falta de conhecimento sobre o assunto e discussão de vários assuntos ao mesmo tempo. Essa pesquisa propõe uma abordagem chamada AidChat com o objetivo de analisar o bate-papo, estruturar o diálogo e recomendar informações complementares a ele. Como abordagem para recomendação de conteúdo foram utilizadas fontes de dados no formato RDF. Estas fontes, disponíveis na Web, fazem parte do cenário da Web 3.0 que tem o objetivo de organizar de forma inteligente informações disponíveis na Web. Estas bases de dados já abrangem diversos contextos e usualmente estão previamente interligadas com outras bases, tornando a informação muito rica. Para avaliar a abordagem proposta foram realizados como experimentos sessões de bate-papos e uma análise qualitativa a partir de respostas de questionários distribuídos aos participantes. Assim foi possível evidenciar que a recomendação de conteúdo complementar ao diálogo, em ambientes de comunicação textual complexos, contribui com a solução de dúvidas tornando o diálogo mais construtivo.

Palavras-chave: Sistemas de Recomendação, Análise do Diálogo, Dados Abertos Interligados

# **Abstract**

LIMA, Bianca Munaro. 2013. 137 f. Dissertação (Mestrado em Informática) - Programa de Pós-Graduação em Informática, Instituto de Matemática, Instituto Tércio Pacitti, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2013.

The popularization of Web 2.0 has led to a considerable increase in the use of textbased computer mediated communication. These, are easy to use and allow synchronous and asynchronous discussions. They make the interaction between people from anywhere in the world possible, quickly revolutionizing the ways of communication. However, this type of communication may also have problems, especially when it only involves text-based communication. This problem occurs because users can only exchange messages in text format, without seeing or talking to each other. Therefore, common problems are: ambiguities in interpretation, understanding problems, difficulties to follow the sequence of the conversation, lack of knowledge on the subject and discussion of various subjects at the same time. This research proposes an approach called AidChat aiming to analyze the chat, structuring dialogue and recommend supplementary information to it. In the approach for content recommendation, linked data sources, in RDF format, has been used. These sources are available on the web and are part of the Web 3.0 scenario that aims to intelligently organize information available on the Web. These databases cover many contexts and usually are previously linked with other databases out of the box, making that information valuable. To evaluate the approach proposed, experiments were performed as chat sessions and as a qualitative analysis of responses from questionnaires distributed to participants. Thus it became clear that content recommendation complements the dialogue, mainly in complex environments, and contributes to the solution of questions making the dialogue more constructive.

Palavras-chave: Recommender Systems, Discourse Analysis, Linked Open Data

# Lista de figuras

Figura 1: Evolução do numero de pessoas com acesso a internet no Brasil	12
Figura 2: Diagrama da nuvem Linking Open Data, por Richard Cyganiak e Anja Jentzsch	16
Figura 3: Exemplo de bate-papo retirado da ferramenta e-Democracia	<b>2</b> 3
Figura 4: Exemplo de decreto retirado da ferramenta LexML	23
Figura 5: Exemplo de bate-papo retirado da ferramenta e-democracia	24
Figura 6: Exemplo de bate-papo retirado da ferramenta e-Democracia	24
Figura 7: Exemplo de bate-papo retirado da ferramenta e-democracia	24
Figura 8: Exemplo de PEC retirado da ferramenta LexML	24
Figura 9: Arquitetura do AidChat	47
Figura 10: Diagrama do módulo de análise do diálogo	53
Figura 11: Modelo de dados para apoio à representação da discussão	54
Figura 12: Exemplo de repetição de palavras em um bate-papo	56
Figura 13: Exemplo de referencia a um usuário em um bate-papo ocorrido no e-	
democracia	57
Figura 14: Exemplo de bate-papo ocorrido na ferramenta e-democracia	57
Figura 15: Exemplo de repetição em bate-papo ocorrido na ferramenta e-democracia	61
Figura 16: Exemplo de bate-papo ocorrido na ferramenta e-democracia	61
Figura 17: Exemplo de referencia a nomes em um bate-papo ocorrido na ferramenta e-	
Democracia	62
Figura 18: Diagrama do módulo de recomendação de conteúdo	66
Figura 19: Exemplo de recomendação no bate-papo implementado	68
Figura 20: Exemplo retirado de bate-papo ocorrido na experimentação	70
Figura 21: Informação sobre a Lei Maria da Penha disponível na Dbpedia	71
Figura 22: Trecho do Comunicógrado gerado na experimentação	73
Figura 23: Seleção de tipos de entidades na DBpedia Spotlight	79
Figura 24: Demonstração das recomendações apresentadas na ferramenta	
implementada	82
Figura 25: Trecho do diálogo realizado na experimentação	87
Figura 26: Trecho do diálogo realizado na experimentação	87
Figura 27: Trecho do diálogo realizado na experimentação	87

Figura 28: Trecho do diálogo realizado na experimentação
Figura 29: Referências entre mensagens no trecho do diálogo realizado na
experimentação89
Figura 30: Análise da quantidade de participantes que marcaram cada resposta94
Figura 31: Análise da quantidade de participantes que marcaram cada resposta94
Figura 32: Análise da quantidade de participantes que marcaram cada resposta95
Figura 33: Análise da quantidade de participantes que marcaram cada resposta95
Figura 34: Análise da quantidade de participantes que marcaram cada resposta96
Figura 35: Análise da quantidade de participantes que marcaram cada resposta97
Figura 36: Análise da quantidade de participantes que marcaram cada resposta98
Figura 37: Análise da quantidade de participantes que marcaram cada resposta98
Figura 38: Análise da quantidade de participantes que marcaram cada resposta99
Figura 39: Análise da quantidade de participantes que marcaram cada resposta100
Figura 40: Análise da quantidade de participantes que marcaram cada resposta101
Figura 41: Análise da quantidade de participantes que marcaram cada resposta101
Figura 42: Análise da quantidade de participantes que marcaram cada resposta102
Figura 43: Análise da quantidade de participantes que marcaram cada resposta103
Figura 44: Análise da quantidade de participantes que marcaram cada resposta104
Figura 45: Análise da quantidade de participantes que marcaram cada resposta104
Figura 46: Análise da quantidade de participantes que marcaram cada resposta105

# Lista de tabelas

Tabela 1: Tabela que representa a distância entre as mensagens de treinamento	listância entre as mensagens de treinamento60
---	---

# Sumário

1	Intro	odução	12
	1.1	Identificação do Problema	17
	1.2	Objetivo	19
	1.3	Abordagem de Solução	19
	1.4	Cenário	21
	1.5	Organização do Texto	25
2	Ref	erencial Teórico	26
	2.1	Comunicação Online	26
	2.2	Análise do Diálogo	29
	2.3	Processamento de Linguagem Natural	36
	2.4	A Iniciativa de Dados Abertos Interligados	38
	2.5	Sistemas de Recomendação	40
	2	.5.1 Recomendação em Ferramentas de Comunicação Textual	41
	2	.5.2 Recomendação de Dados Abertos Interligados	42
3	Aid	Chat – Sistema de Recomendação para Bate-papos	44
	3.1	Definição de Requisitos	44
	3.2	Abordagem Proposta	45
	3.3	Arquitetura de Apoio à Abordagem	46
	3	.3.1 Módulo de Pré-processamento de Mensagens	47
	3	.3.2 Módulo de Análise do Diálogo	52
	3	.3.3 Módulo de Recomendação de Conteúdo	65
	3.4	Problemas Encontrados	77
4	Ava	aliação	.80
	4.1	Implementação no Cenário Escolhido	80
	4.2	Primeira Experimentação	82
	4	.2.1 Avaliação de Perguntas Abertas	83
	4.3	Segunda Experimentação	85
	4	.3.1 Avaliação do Diálogo	86
	4	.3.2 Avaliação de Perguntas Abertas	89
	4.4	Avaliação de perguntas Fechadas	93
5	Cor	nclusão1	.07
	5.1	Contribuições Principais	.08

	5.2 Limitações	.09
	5.3 Trabalhos Futuros	09
6	Referências1	.13
Α	pêndices 117	
	APÊNDICE A – respostas de perguntas ABERTAS do questionário utilizado nas du experimentações	
	Quais problemas (que geralmente ocorrem em chats com múltiplos usuários) vo percebeu que a ferramenta não contribuiu para solucionar?	
	Quais problemas (que geralmente ocorrem em chats com múltiplos usuários) vo percebeu que a ferramenta contribuiu para solucionar?	
	Gostaria de fazer mais algum comentário? Sugestões?1	19
	APÊNDICE B – diálogo e recomendações geradas a partir da segunda experimentação 1	22

# 1 Introdução

O número de usuários que utilizam a Web ativamente, no mundo inteiro, está crescendo significativamente. No Brasil, de acordo com o IBOPE Nielsen Online<sup>1</sup>, o acesso à Internet atingiu 77,8 milhões de pessoas no segundo trimestre de 2011, representando um aumento de 20% em relação ao segundo trimestre de 2009. Essa expansão está associada à diminuição do custo de acesso, ao crescimento do número de computadores em residências e ao surgimento de diversas tecnologias e novos conceitos que trazem facilidades para o cotidiano da população.

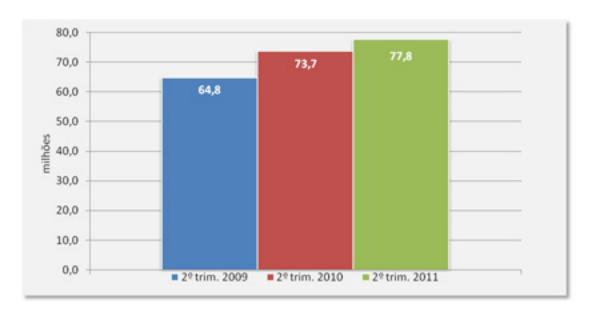


Figura 1: Evolução do número de pessoas com acesso a internet no Brasil.

Hoje em dia, através da Internet, é possível fazer compras, agendar viagens, fazer reservas em hotéis, fazer ligações, participar de vídeo conferências e conversar com amigos, tudo isso em uma velocidade muito grande. Para que tantas facilidades sejam oferecidas aos usuários, a Web sofreu grandes mudanças ao longo dos anos. Desde o período de sua criação, além da natural evolução das tecnologias associadas, seus objetivos foram se expandindo, ocorrendo mudanças na natureza de sua utilização.

A chamada Web 1.0, considerada a primeira geração da Web, permitia a leitura e a criação de ligações entre documentos constituídos basicamente por páginas estáticas. Nessa

<sup>&</sup>lt;sup>1</sup> Total de pessoas com acesso à internet atinge 77,8 milhões.Disponível em: <a href="http://www.ibope.com.br/pt-br/noticias/Paginas/Total">http://www.ibope.com.br/pt-br/noticias/Paginas/Total</a> de pessoas com acesso à internet atinge 77,8

época, existia pouca interação e colaboração entre os usuários, pois somente o responsável pelo domínio em questão podia alterar ou adicionar conteúdo ao site. A ideia dessas páginas era apenas divulgar uma informação e o papel do usuário se concentrava no consumo do conteúdo publicado.

Em 2004, a empresa americana O'Reilly Media passa a associar a expressão Web 2.0 ao conceito de uma Internet colaborativa, onde usuários interagem e geram conteúdo, graças ao surgimento de ferramentas intuitivas e de fácil utilização. Dentre elas podemos citar as wikis, blogs, redes sociais, entre outras ferramentas em que o foco é a interação. De acordo com Oreilly (2007), o sucesso dessas ferramentas se deve ao fato de que elas aproveitam a inteligência coletiva de forma que quanto mais são utilizadas, melhores ficam. Um exemplo é o eBay, que cresce de acordo com a atividade dos usuários. Dessa maneira, a grande massa de compradores e vendedores torna a ferramenta mais atraente que qualquer novo serviço semelhante. Devido a isso, novos vídeos, notícias em blogs, comentários e avaliações são criados a todo o momento, contribuindo para o aumento da quantidade de informação disponível na Web.

No entanto, as informações geradas pelas comunidades de cada uma das ferramentas da Web 2.0 se encontram fragmentadas (BIZER 2009), além de frequentemente possuírem inconsistências e baixa qualidade. Isso acontece devido ao sistema livre de colaboração nessas ferramentas, a necessidade de disseminar informação com velocidade e por não serem usados identificadores únicos para representação de dados.

Cada ferramenta organiza e gerencia o conteúdo gerado e por ela mantido, havendo cooperação e integração geralmente apenas entre ferramentas de mesmo fornecedor. Mesmo neste último caso, essa integração ainda é insipiente. Como exemplo desta fragmentação, o eBay<sup>2</sup> tem uma base enorme de indivíduos vendendo produtos, o LinkedIn<sup>3</sup>

O LinkedIn é uma rede social focada em negócios e relações entre profissionais e empresas. Disponível em: <a href="http://br.linkedin.com">http://br.linkedin.com</a>. 12/01/2014

eBay é o nome de uma empresa de comércio eletrônico fundada nos Estados Unidos. Atualmente é o maior site do mundo para a venda e compra de bens. Disponível em: <a href="http://www.ebay.com">http://www.ebay.com</a>>. 12/01/2014

tem informações sobre pessoas, seus empregos e as relações entre elas, o Google<sup>4</sup> indexa o conteúdo de sites, o Facebook<sup>5</sup> tem uma base de usuários, empresas e opiniões sobre elas. Essas aplicações têm muitos pontos em comum: elas tentam relacionar informação com necessidade, mas todas com sua própria base e escopo de atuação, limitando-se à exploração de seus próprios dados. No entanto, é evidente o potencial da exploração conjunta dessas informações, conectando-se itens de dados associados aos mesmos referentes no mundo real. Atualmente, já existem algumas iniciativas de integração das informações. Como exemplo, diversas aplicações permitem que o usuário se registre utilizando a conta do Google ou do Facebook.

Um estudo da Cisco<sup>6</sup> divulgou que a Internet crescerá quatro vezes de 2011 a 2016. O estudo estima um crescimento de consumidores móveis de 3,7 bilhões em 2011 para 4,5 bilhões em 2016 e de 1,7 bilhão de internautas com acesso residencial a Internet para 2,3 bilhões em 2016. Hoje em dia, são gerados<sup>7</sup> 50 milhões de tweets por dia, 24 petabytes de dados processados pelo Google por dia e 20 horas de vídeos adicionados ao Youtube por minuto. Portanto, com o crescimento da utilização da Internet e, consequentemente, das informações disponíveis, é necessário que novas técnicas sejam usadas para que a utilização dessas informações seja feita de forma eficiente.

Outro desafio associado à escalabilidade do tratamento dessas informações é que, originalmente, estes ambientes interativos tinham por alvo a interpretação e o entendimento por humanos. Os navegadores da Web apenas intermediam a apresentação de formatos como HTML (*HyperText Markup Language*), DOC (*Document*) e PDF (*Portable Document Format*). Com isso, há considerável dificuldade para que um sistema processe e

-

<sup>&</sup>lt;sup>4</sup> O Google é uma empresa multinacional de serviços online e software dos Estados Unidos. O Google Search é um serviço da empresa onde é possível fazer pesquisas na internet sobre qualquer tipo de assunto ou conteúdo. Disponível em: <a href="http://www.google.com">http://www.google.com</a>>. 12/01/2014

<sup>&</sup>lt;sup>5</sup> Facebook é um site e serviço de rede social que foi lançada em 4 de fevereiro de 2004, operado e de propriedade privada da Facebook Inc. Disponível em: <a href="http://www.facebook.com">http://www.facebook.com</a>>. 12/01/2014

<sup>&</sup>lt;sup>6</sup> Cisco VNI estima que a Internet crescerá quatro vezes nos próximos quatro anos. Disponível em: <a href="http://globalnewsroom.cisco.com/easyir/BR/pt/local/press-release/Cisco-VNI-estima-que-a-Internet-crescera-quatro-vezes-nos-proximos-quatro-anos--893289.html">http://globalnewsroom.cisco.com/easyir/BR/pt/local/press-release/Cisco-VNI-estima-que-a-Internet-crescera-quatro-vezes-nos-proximos-quatro-anos--893289.html</a>>. 12/11/2012

<sup>&</sup>lt;sup>7</sup> The World of Data We're Creating on the Internet. <a href="http://www.good.is/posts/the-world-of-data-we-re-creating-on-the-internet/">http://www.good.is/posts/the-world-of-data-we-re-creating-on-the-internet/</a>. 11/11/2012

interprete essas informações de forma automática. Para fazer frente a esse desafio, surgiu o conceito de Web 3.0 ou Web Semântica, que tem como objetivo organizar, de forma mais inteligente, o conteúdo presente na Web, representando os dados de forma que os computadores também possam interpretá-los, associando descritores aos dados e às suas ligações, baseados em padrões gerais de representação e em vocabulários definidos para domínios específicos (LASSILA; HENDLER 2007). Com isso, torna-se possível relacionar dados de diversos contextos e construir aplicações que os apresentem de forma integrada, fazendo com que a informação tenha um valor muito maior devido aos relacionamentos com outros dados.

Considerada por muitos como a abordagem que tornou realidade a Web Semântica<sup>8</sup>, surge a proposta de dados abertos interligados (LOD - Linked Open Data ), explorando de forma bastante criativa tecnologias e padrões já existentes. Uma delas é o RDF (Resource Description Framework), padronizado pelo W3C (LASSILA; SWICK 1999) e utilizado agora como modelo de representação para a publicação de dados na Web. O RDF, apresentado originalmente como um padrão para representação e interoperabilidade de metadados, mostra-se agora como uma forma de representação simples e flexível, permitindo a descrição de recursos, suas propriedades e relações na forma de triplas do tipo (recurso-propriedade-valor). Cada propriedade pode ser anotada, referenciando vocabulários ou ontologias existentes criadas para aquele domínio ou aplicação.

Em 2007, existiam apenas 12 bases publicadas nesse formato e em 2011 esse número já crescia para 295<sup>9</sup>, incluindo bancos de dados de diferentes domínios. Um exemplo importante é a DBPedia (LEHMANN et al. 2013), gerada a partir de dados estruturados disponíveis na Wikipedia, que se tornou uma espécie de *data hub*, servindo de referência para a interligação de diversas outras bases por se tratar de uma base de dados com conteúdo muito abrangente.

<sup>&</sup>lt;sup>8</sup> Linked Data. Disponível em: <a href="http://www.w3.org/standards/semanticweb/data">http://www.w3.org/standards/semanticweb/data</a>. 29/04/2012

<sup>&</sup>lt;sup>9</sup> Linked Data Disponível em: <a href="http://linkeddata.org/">http://linkeddata.org/</a>>.14/04/2012

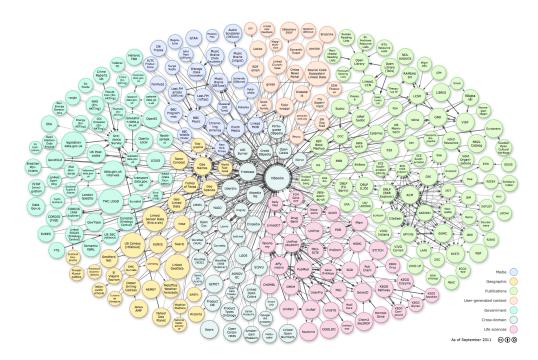


Figura 2: Diagrama da nuvem Linking Open Data, por Richard Cyganiak e Anja Jentzsch.

Fonte: http://lod-cloud.net/

Com a proliferação de bases publicadas em LOD (também chamada de Web de Dados – WoD), surgem cada vez mais aplicações que exploram seus dados de forma conjunta, extraindo novas informações a partir daquelas já publicadas. Dessa maneira, torna-se possível solucionar diversos problemas que ocorrem durante a utilização da Web. Dentre esses problemas estão a necessidade de consulta a diversos sites para um mesmo fim, a exemplo de agências de viagens, companhias aéreas e sites com publicações de artigos, ter que fazer um mesmo cadastro em diversos sites, como em redes sociais e lojas online, e ainda a dificuldade para encontrar informações em ferramentas de busca.

O surgimento dessa nova forma de organização da informação também permite analisar e estruturar, de maneira semântica, dados presentes em aplicações da Web 2.0, como redes sociais. Um exemplo seria a análise das opiniões sobre uma eleição no Twitter (TUMASJAN et al. 2010). Posteriormente, esses dados poderiam ser gravados em uma base de dados de triplas e associados com uma base de dados sobre políticos. Dessa forma, poderíamos registrar a evolução das opiniões sobre um político no decorrer das eleições. Esse tipo de análise pode agregar dados muito construtivos para a sociedade, e, portanto, seria importante analisar, salvar e integrar esses dados com outras bases relacionadas. Esse é só um exemplo do potencial da Web 3.0. É necessário que, cada vez mais, dados

interessantes sejam publicados nesse formato e que sejam desenvolvidas aplicações inteligentes a partir deles. Assim, é possível diminuir o tempo para encontrar uma informação e sua relação com outros dados.

# 1.1 Identificação do Problema

As possibilidades de evolução das aplicações que os conceitos de LOD podem trazer para a Web e os problemas nela vivenciados geram diversas possibilidades no contexto de contribuições para a qualidade e rapidez da navegação na Web, explorando de maneira mais abrangente as informações existentes.

Atualmente, um dos maiores problemas encontrados ao navegar na Web é a quantidade crescente de dados nela disponíveis, dificultando a busca por um conteúdo específico e o aproveitamento de todo o potencial que esses dados podem trazer se utilizados em conjunto. As informações provenientes de blogs, wikis e diversas redes sociais, também contêm conteúdos interessantes que muitas vezes acabam se perdendo. Assim, cada vez mais, existe a necessidade de construir ferramentas que consumam esses dados de forma inteligente. A organização da informação está sendo realizada a partir da publicação da mesma, seguindo as regras definidas pela Web 3.0, permitindo que as informações sejam utilizadas em conjunto e consumidas de maneira simples pelas aplicações. Nesse cenário, também é importante que sejam desenvolvidas soluções que consumam esses dados, melhorando a experiência de navegação na Web.

Dessa forma, esse trabalho propõe uma abordagem que contribui para a solução de um problema de comunicação muito comum vivenciado pelos usuários da Web diariamente durante bate-papos online. A abordagem tem como foco o problema de falta de conhecimento do assunto e de solução de conceitos ambíguos, utilizando dados disponíveis de acordo com os conceitos de LOD.

A motivação para essa visão está relacionada com a popularização da Web 2.0 que provocou um considerável aumento na utilização dos meios de comunicação textuais. Esses são de fácil uso e permitem discussões assíncronas e síncronas, em tempo real. A comunicação textual é muito utilizada por diversos grupos de usuários para trocar informações. Alguns exemplos desse uso são: a comunicação em redes sociais, o ensino a distância, a comunicação no ambiente de trabalho, aplicações de correio eletrônico, entre

outros.

A comunicação mediada por computador tem como vantagens o seu alcance, a velocidade, o baixo custo, a possibilidade de ser síncrona ou assíncrona, as facilidades de apoio para tratamento e acesso às informações trocadas, além das possibilidades de análise sobre essas interações.

Atualmente, esse meio de comunicação é empregado em grande escala diariamente por grande parte das pessoas que utilizam a Web. Ele permite a interação entre pessoas de qualquer parte do mundo rapidamente, revolucionando as formas de comunicação. Ele é usado tanto em meios de comunicação que permitem a comunicação assíncrona, como o correio eletrônico, redes sociais, wikis, jogos online, quanto em meios de comunicação síncronos como os bate-papos e conferências por vídeo.

No entanto, esse tipo de comunicação pode também apresentar dificuldades, principalmente quando envolve apenas comunicação textual. Nela, os usuários utilizam apenas a troca de mensagens na forma de texto, não podendo se ver ou falar. O problema é que muitas vezes os usuários podem não expressar suas ideias com clareza, ou mesmo as palavras podem sofrer interpretações errôneas, por terem significado ambíguo ou por estarem apresentadas ou consideradas em contexto distinto daquele intencionado pelo emissor. Outro problema ocorre quando o participante não tem conhecimento sobre o assunto ou sobre determinados termos.

Portanto, é comum que existam ambiguidades na interpretação, dificuldades de entendimento, dificuldades de acompanhar a conversa (PIMENTEL 2002) e discussão de vários assuntos ao mesmo tempo. Os problemas relatados ocasionam a perda de tempo com buscas, perguntas, esclarecimentos e desentendimentos (HORIGUCHI et al. 2009). Quando um usuário tem uma dúvida, ele faz uma pergunta, uma busca na Web ou procura em mensagens anteriores onde se perdeu.

As dificuldades relatadas acima ocorrem tanto em bate-papos<sup>10</sup> com dois participantes quanto em diálogos com um grupo de participantes. O foco desse trabalho são os diálogos com três ou mais pessoas. Estes envolvem desafios que não ocorrem com

 $<sup>^{10}</sup>$  O termo coloquial do português foi usado para designar o termo em inglês,  ${\it chat}$ 

apenas dois usuários. Devido ao número de pessoas, o diálogo pode se tornar mais confuso, causando mais dificuldades e necessitando mais ainda de medidas que facilitem a comunicação.

Quando a conversa acontece entre amigos, em redes sociais, esses problemas não causam grandes impactos, porém, fóruns e bate-papos também são utilizados para a discussão de assuntos mais importantes. Nos ambientes de educação, esses problemas podem fazer com que um aluno não consiga se comunicar ou não entenda parte da matéria, prejudicando seu aprendizado. Já nas comunidades que utilizam a comunicação textual para negociação de produtos, os desentendimentos podem atrapalhar uma venda ou promover propaganda negativa contra a empresa. Portanto, esse trabalho pretende contribuir para a solução de problemas de comunicação que acontecem em ambientes em que o contexto é bem definido, o conteúdo do diálogo é mais sério e os impactos causados por esses problemas geralmente são maiores.

## 1.2 Objetivo

Esse trabalho tem o objetivo de propor uma abordagem para a análise e solução de problemas que, geralmente, ocorrem em bate-papos realizados em ferramentas de comunicação textual. Os problemas encontrados ocorrem principalmente nos diálogos em que grupos de usuários participam.

Como revisado no capítulo anterior diversos problemas de comunicação podem ocorrer em bate-papos. Essa pesquisa efetua a revisão destes, além dos trabalhos relacionados e propostas de solução existentes. A partir dessa análise, uma abordagem para mitigar esses problemas é proposta.

Para tornar possível a avaliação da abordagem, ela foi implementada em uma ferramenta de comunicação textual já existente. Assim, posteriormente, foi possível testar a abordagem em sessões de bate-papos e avaliar a contribuição para o diálogo.

### 1.3 Abordagem de Solução

A abordagem proposta visa analisar as mensagens de um bate-papo, descobrir os pontos em que ocorreram problemas de comunicação e recomendar informações interessantes que colaborem construtivamente com a discussão. A motivação para a

recomendação de conteúdo está nos problemas de entendimento ocasionados por palavras ambíguas ou conceitos que o usuário desconhece. A seguir, nos próximos capítulos, pode-se perceber que existe uma literatura extensa para a solução dos problemas de entendimento dificuldade de acompanhar 0 encadeamento do diálogo. pela Portanto, complementarmente, essa pesquisa tem o objetivo de contribuir para minorar outros problemas que ocorrem em bate-papos online recomendando informações complementares. Assim, é possível que dúvidas sejam solucionadas e usuários entendam melhor o assunto sendo discutido proporcionando uma conversa mais rápida e eficiente.

Portanto, para a concretização dessa ideia, inicialmente é necessário entender como funciona o processo de comunicação para que seja possível identificar o surgimento de dúvidas e desentendimentos. A partir dessa análise, para complementar a troca de informações, dados presentes em bases de dados abertas podem ser recomendados, de acordo com a interação em andamento. Desse modo, novos conhecimentos são gerados a partir da agregação de novas informações, solucionando dúvidas rapidamente, no próprio ambiente da ferramenta, sem a necessidade da troca de mensagens adicionais explicativas.

Porém, a maior dificuldade dessa abordagem é descobrir quais dados podem agregar conteúdo que contribua com a comunicação, fazendo com que os participantes possam adquirir conhecimento e esclarecer conceitos ambíguos. Para isso é necessário analisar as mensagens e identificar relações entre mensagens e momentos em que usuários estão com dúvida, tendo como objetivo identificar usuários que não estejam conseguindo acompanhar a discussão. Utilizando os dados recuperados nessa análise, é possível descobrir informações que possam auxiliar na comunicação para que ocorram menos dúvidas, conflitos e usuários confusos. Em ambientes em que o diálogo tenha focos mais específicos e complexos, a necessidade por informação complementar é ainda maior.

Atualmente, existem muitas comunidades virtuais, focadas em diversos contextos, que se utilizam das redes sociais para interação e ações de seu interesse. Um típico exemplo é o e-Democracia<sup>11</sup>, um portal desenvolvido pela Câmara dos Deputados que tem o objetivo de incentivar a participação da sociedade no debate de temas importantes para o país. Dessa

e-Democracia – Participação virtual, cidadania real Disponível em <a href="http://edemocracia.camara.gov.br/">http://edemocracia.camara.gov.br/</a>>. 04/04/2013

maneira, são disponibilizados espaços para a discussão de projetos de leis entre a população. Com isso, a participação da população aumenta, tornando a criação de leis mais transparente e incentivando participantes a discutir, dar ideias e interagir com deputados. O problema é que a legislação é muito extensa e esse conhecimento é necessário para a discussão dos novos projetos de leis, visto que é importante entender as leis que já existem no contexto sendo discutido. Portanto, nesse cenário, a utilização de bases de dados que agreguem informações sobre leis e seus contextos poderiam facilitar o diálogo.

Portanto, a complementação de informação pode diminuir o tempo que os usuários necessitam para fazer buscas na Web e perguntas, pois eles passam a adquirir conhecimento diretamente via a ferramenta. Isso faz com que o conhecimento seja mais bem distribuído, fazendo com que todos os usuários tenham acesso a uma informação em vez de cada um buscar esses dados em lugares diferentes.

Essa é uma motivação para o uso de dados abertos, pois o objetivo da Web Semântica é justamente manter o conhecimento centralizado e interligado. Além disso, a escolha de bases de dados abertas, no formato RDF, permite que sejam utilizadas diversas bases e que muitas delas já estejam previamente interligadas, abrangendo diversos contextos. De acordo com a evolução e crescimento da Web 3.0 e as bases de dados disponíveis, a solução proposta pode se tornar mais poderosa abrangendo uma grande quantidade de assuntos.

#### 1.4 Cenário

Nessa pesquisa, foi necessário escolher um cenário específico em que problemas de comunicação textual ocorrem com frequência para que a abordagem proposta pudesse ser avaliada. Portanto, foi escolhido um cenário em que os bate-papos são focados em assuntos sérios e complexos fazendo com que os problemas de comunicação tenham um grande impacto sobre o objetivo do diálogo.

A escolha de um cenário específico também contribui para a validação da proposta, mostrando de forma mais clara os objetivos e benefícios alcançados pela abordagem proposta. Portanto, ela será aplicada em um cenário em que existe uma grande necessidade de melhoria do diálogo e onde a proposta pode trazer grandes benefícios.

O contexto escolhido para aplicação é a legislação brasileira. O e-Democracia é um

portal desenvolvido pela Câmara dos Deputados que permite que a população participe de debates sobre leis, projetos de leis e apresente suas ideias e contribuições. O envolvimento dos cidadãos nesses debates permite que haja mais transparência e que os deputados possam entender melhor as necessidades e anseios da população. O portal conta com diversas comunidades geralmente relacionadas a projetos de leis, que abrangem fóruns, uma biblioteca virtual com documentos relacionados, uma wiki para discussão do projeto de lei, um espaço para a participação em assembleias legislativas e bate-papos entre membros das comunidades e deputados envolvidos com o tema.

Esse cenário foi escolhido, pois, nesse contexto, é necessário que os usuários tenham o conhecimento de dados de grande volume e complexidade. Atualmente, o senado brasileiro disponibiliza diversos dados abertos. Entre eles dados de parlamentares, do plenário, legislativos, jurídicos e de projetos. Portanto, podem ser utilizados, por exemplo, dados da base do LexML<sup>12</sup>, constituída de dados legislativos e jurídicos, para melhorar a qualidade do bate-papo.

O LexML é uma iniciativa conjunta de diversos órgãos, liderada pelo Senado Federal. Seu objetivo é identificar e estruturar informações legislativas e jurídicas e disponibiliza-las utilizando padrões de dados abertos. A estruturação dessa informação permite a criação de ligações entre diversas manifestações de um texto nos diversos órgãos e correlacionar citações, por exemplo. O LexML reúne leis, decretos, acórdãos, súmulas, projetos de leis entre outros documentos das esferas federal, estadual e municipal dos Poderes Executivo, Legislativo e Judiciário de todo o Brasil. Os dados estão no formato XML e um modelo de referencia foi definido para modelagem conceitual desses dados.

Portanto, devido à complexidade e à quantidade de dados legislativos, podem ocorrer diversos problemas no diálogo em bate-papos do e-democracia. Assim, analisando as mensagens do bate-papo é possível descobrir palavras que representem os assuntos mais discutidos e ações que representem dúvidas dos participantes da discussão. Utilizando esses dados, em conjunto, podem ser encontrados conteúdos relacionados que sejam interessantes aos usuários. Essa etapa também permite que muitos outros dados sejam

 $<sup>^{12}</sup>$  LexML – Rede de informação legislativa e jurídica. Disponível em: < http://www.lexml.gov.br/ >. 17/09/2012

extraídos das mensagens possibilitando que outras contribuições sejam feitas a comunicação.

O e-Democracia armazena os bate-papos realizados durante assembléias públicas e os disponibiliza para download nos formatos XML e HTML. Assim, é possível analisar os diálogos disponibilizados pela ferramenta e avaliar os problemas de comunicação ocorridos.

A figura 3 apresenta uma evidência de falta de conhecimento, do usuário, em relação ao assunto sendo discutido:

```
Você acha que as leis atuais que combatem o tráfico de pessoas são adequadas e suficientes?

Elas estão sendo bem aplicadas?
```

Não tenho dados, mas creio que a maioria dos casos de tráfico de pessoas seja para a prostituição, como não temos uma regulamentação dessa atividade aqui no Brasil fica difícil viabilizar esse registro formal das vagas.

Figura 3: Exemplo de bate-papo retirado da ferramenta e-Democracia

Esse diálogo foi retirado do fórum de uma comunidade que discute a legislação que combate o tráfico de pessoas no Brasil. Para estimular a troca de ideias entre os participantes, a equipe pergunta se as leis existentes, nesse contexto, são adequadas. A figura 3 apresenta a resposta do usuário, manisfestando que não tem conhecimento das leis existentes nesse contexto e apresentando sua opinião sobre o assunto. Portanto, caso o usuário tivesse conhecimento das leis existentes, ele poderia apresentar uma contribuição muito maior e sugerir melhorias. Fazendo uma busca no LexML por tráfico de pessoas, podemos encontrar leis, decretos e projetos de leis existentes na legislação atual. A figura 4 destaca uma das leis encontradas a partir da busca no LexML.

```
Localidade Brasil

Autoridade Federal

Título Decreto nº 7.901, de 04 de fevereiro de 2013

Data 04/02/2013

Ementa INSTITUI A COORDENAÇÃO TRIPARTITE DA POLITICA NACIONAL DE ENFRENTAMENTO AO TRAFICO DE PESSOAS E O COMITÊ NACIONAL DE ENFRENTAMENTO AO TRAFICO DE PESSOAS - CONATRAP.

URN urn:lex:br:federal:decreto:2013-02-04;7901

Assuntos {CRIAÇÃO, INSTITUIÇÃO}, COMPOSIÇÃO, {COMPETENCIA, ATRIBUIÇÃO}, COMISSÃO... mais
```

Figura 4: Exemplo de decreto retirado da ferramenta LexML

Na figura 5, dando continuidade à discussão, o mesmo usuário comenta que deveria

ter uma legislação para intercâmbio estudantil e demonstra não estar muito ciente das leis existentes para esse contexto também.

```
RE: Legislação
16/03/13 17:51 em resposta a Equipe e-Democracia.

Seria bom também haver uma legislação específica para intercâmbio estudantil, até onde eu sei, isso não é uma atividade regulamentada, é muito fácil criar uma empresa de fachada.
```

Figura 5: Exemplo de bate-papo retirado da ferramenta e-democracia

Outro exemplo (figura 6), no contexto da reforma política mostra que o usuário não tem conhecimento das propostas feitas para a reforma. Porém, esses dados estão disponíveis na biblioteca virtual do e-Democracia. Portanto seria interessante que essas informações também fossem apresentadas no bate-papo, fazendo com que o usuário pudesse acessá-las no decorrer do diálogo tornando a discussão mais produtiva.

```
(05/11/2013 - 10:40:45) - FELIPE fala : queria o texto comleto dessa reforma. Alquém ai tem o link? (05/11/2013 - 10:50:13) - Liane Lira fala : Não há um texto único para a reforma (05/11/2013 - 10:50:31) - Liane Lira fala : as propostas estão em projetos de lei diferentes (05/11/2013 - 10:50:41) - Liane Lira fala : dá uma olhada aqui, FELIPE: http://edemocracia.camara.gov.br/web/reforma-politica/proposicoes#.UnjpZ1RDs2N (05/11/2013 - 10:50:59) - Liane Lira fala : (na Biblioteca Virtural) (05/11/2013 - 10:52:10) - FELIPE fala : LIANE, valeu, estou com um olho aqui e outro no serviço haha
```

Figura 6: Exemplo de bate-papo retirado da ferramenta e-Democracia

Ainda nesse contexto é apresentada na figura 7 a dúvida de um participante que não conhece o texto de uma proposta de emenda à constituição (PEC).

```
(15/08/2013 - 10:36:42) - David Joade Lima Melo fala para Tiago Costa: achei bastante interessante o seu questionamento!! É bom lembrar que tem umas PECs tramitando na Câmara que vão nessa direção! Uma delas é a PEC 434/2009 de autoria do deputado Viera da Cunha... é uma proposta mto interessante!! (15/08/2013 - 10:37:19) - Tiago Costa fala para David Joade Lima Melo: Em suma o que diz essa pec?
```

Figura 7: Exemplo de bate-papo retirado da ferramenta e-democracia

A informação representada na figura 8 pode ser encontrada na base de dados do LexML:

```
Autoridade Câmara dos Deputados

Título PEC 434/2009

Data 12/11/2009

Ementa Dá nova redação ao art. 101 da Constituição Federal, para alterar a forma e requisitos pessoais de investidura no Supremo Tribunal Federal.
```

Figura 8: Exemplo de PEC retirado da ferramenta LexML

Portanto, a análise dos diálogos realizados no portal do e-Democracia mostra que muitas vezes a população em geral não tem o conhecimento das leis existentes. O portal disponibiliza uma biblioteca com informações da legislação vigente no tema da comunidade

em questão. Porém, percebemos que muitas vezes os usuários não entram na biblioteca para estudar as leis antes de participar de um bate-papo. Outro problema é que não é possível apresentar todas as leis de temas relacionados na biblioteca, pois a legislação é muito extensa e existem inúmeras possibilidades. No caso do exemplo apresentado no tema do tráfico de pessoas, o usuário relaciona o tema com a legislação para intercâmbio estudantil. Assim, a recomendação em tempo real tem uma grande vantagem por permitir a busca da legislação a partir do diálogo sendo realizado.

## 1.5 Organização do Texto

A organização do texto deste trabalho é detalhada a seguir. Este capítulo inicial tem o objetivo de apresentar uma introdução à pesquisa, incluindo o problema abordado, as motivações e objetivos, além de descrever o cenário em que ela foi aplicada. O segundo capítulo apresenta o referencial teórico para desenvolver a abordagem proposta e trabalhos relacionados usados como referencia para propor essa contribuição.

O capítulo 3 descreve a abordagem e os detalhes da sua arquitetura e implementação, bem como os testes realizados durante a implementação. No quarto capítulo a avaliação da abordagem é discutida destacando os detalhes da metodologia aplicada, as experimentações realizadas e os resultados encontrados. Por fim, no quinto capítulo, a conclusão é apresentada destacando contribuições, limitações e trabalhos futuros.

# 2 Referencial Teórico

Esse capítulo apresenta a revisão de um conjunto de trabalhos que trazem contribuições ao tema de análise da comunição. Além disso, também foram revisadas áreas em que as técnicas foram necessárias para a implementação da abordagem proposta. São elas, o processamento de linguagem natural, usado na etapa de análise do diálogo, a recomendação de conteúdo e a iniciativa de dados abertos interligados, pois bases de dados dessa iniciativa foram utilizadas como fonte de informação na abordagem de recomendação de conteúdo.

### 2.1 Comunicação Online

A Internet tem revolucionado o mundo da comunicação como nenhuma outra invenção foi capaz de fazer antes. A Internet é, de uma vez e ao mesmo tempo, um mecanismo de disseminação da informação e divulgação mundial e um meio para colaboração e interação entre indivíduos e seus computadores, independentemente de suas localizações geográficas. Portanto, desde seu inicio, o objetivo principal da Web sempre foi facilitar a comunicação possibilitando a comunicação entre usuários ao redor do mundo rapidamente.

A comunicação através da Internet possibilita aos usuários ler, escrever e interagir através de computadores conectados a Web. Essa comunicação, dependendo do meio utilizado, pode ser síncrona, assíncrona e alguns meios permitem tanto a interação síncrona quanto assíncrona. A primeira é aquela que se dá em tempo real, ou seja, os usuários que estão envolvidos na troca de mensagens têm acesso imediato às respostas e reações do outro. Já na segunda, a troca de mensagens é intercalada por um período de tempo que desconfigura o caráter de instantaneidade da comunicação síncrona. Por não pressupor que os usuários estejam disponíveis em um mesmo momento, esses softwares armazenam a mensagem para que o usuário possa respondê-la posteriormente.

No início da sua criação, a Internet permitia apenas a leitura e escrita de documentos como forma de comunicação, portanto era permitida apenas a interação assíncrona. Posteriormente, inúmeros meios foram sendo criados para facilitar a troca de informação.

Atualmente, existem diferentes meios de comunicação disponíveis que permitem interações síncronas e assíncronas em ambientes que proporcionam desde a troca de mensagens de texto até a comunicação por áudio e vídeo.

Os meios de produção e divulgação de informação, em que antes apenas uma pequena parcela da população tinha acesso, hoje estão disponíveis para todas as pessoas com acesso a Internet. Com a popularização de celulares, *smartphones*, *tablets* e *notebooks* cada vez mais baratos e mais fáceis de usar, as pessoas conectadas a Web têm produzido e divulgado blogs, vídeos, músicas, notícias, opiniões entre outros.

Portanto, com o aumento da divulgação dos meios de comunicação na Web, surgiu a oportunidade de utilizá-los em diversos segmentos como entretenimento, comércio e educação. Assim, esses meios tornam as comunicações interpessoais mais ágeis e simples, utilizando fóruns, grupos de discussão e redes sociais.

Na educação, as novas tecnologias permitem que sejam criados ambientes ricos em possibilidades de aprendizagem. Assim, as pessoas interessadas e motivadas podem adquirir conhecimento de diversas áreas sem a necessidade de se locomover. Com a popularização da Internet, isso torna o ensino acessível a uma grande parcela da população que está longe dos grandes centros ou não tem tempo para frequentar aulas presenciais.

O termo Intranet surgiu no início de 1994, nos EUA, no vocabulário da informática e na administração das empresas. Ele designa uma rede de computadores privada que funciona dentro dos limites bem definidos de uma organização, podendo ser acessada apenas por usuários internos. Com o seu surgimento, as empresas passaram a utilizar meios de comunicação como o e-mail, para enviar uma mensagem para seus empregados, e batepapos, como meio de comunicação entre empregados localizados em locais afastados.

Portanto, é possível perceber que a comunicação mediada por computador trouxe diversas vantagens para diferentes grupos de usuários. No entanto, ela também pode causar conflitos, principalmente no meio de comunicação textual. Muitas vezes os usuários não sabem se expressar, não escrevem claramente suas ideias e têm culturas diferentes, causando problemas na comunicação.

Os bate-papos são uma forma de comunicação mediada que podem ser utilizados de duas formas: entre dois usuários ou entre múltiplos usuários. Essas duas formas de

comunicação possuem particularidades diferentes. Na comunicação entre apenas dois participantes, é mais fácil trocar mensagens e seguir uma conversa parecida com o diálogo face-a-face. Isso acontece, pois menos assuntos são discutidos ao mesmo tempo e é mais fácil respeitar as regras que utilizamos normalmente para indicar que é a vez da outra pessoa falar, como o silencio, uma pergunta, entre outros. Porém, em um grupo de usuários, mesmo que os elementos da conversação falada continuem sendo usados (HILGERT 2000), é mais difícil organizar o diálogo, causando maiores problemas para o entendimento da discussão.

Nos bate-papos é muito comum o uso de abreviações, gírias, siglas e caracteres diferentes. Muitas dessas siglas são utilizadas para abreviar palavras e frases ou aproximar a escrita da fonética. Isso faz com que o texto fique ainda mais confuso causando dificuldade para o seu entendimento.

Além disso, geralmente, o texto gerado em uma sessão de bate-papo segue um formato não linear, pois nem sempre as mensagens estão associadas com a mensagem que as antecede. Isso acontece porque muitos assuntos são discutidos ao mesmo tempo e é comum ocorrerem referências a mensagens anteriores. Esse fato é muito influenciado pelo número de participantes, pois existe uma chance maior de que um deles introduza uma mensagem no meio de duas mensagens relacionadas. Essa distancia entre as mensagens pode dificultar o acompanhamento da discussão. Assim, os usuários acabam se sentindo "perdidos" e ocorre a perda de co-texto (PIMENTEL et al, 2003). O termo perda de co-texto, segundo Pimentel, foi elaborado para representar o momento em que o leitor não consegue estabelecer o encadeamento da conversação.

A manifestação textual da perda de co-texto permite que a identificação do fenômeno seja mais simples, porém a perda também pode ocorrer sem que haja a manifestação do usuário. Existem fatores que podem potencializar a perda de co-texto. A distância entre as mensagens é um fator que dificulta o acompanhamento da discussão. Outro fator é a quantidade de assuntos sendo discutidos em paralelo. As características cognitivas do interlocutor, como memória, atenção e interesse e as características do grupo, como número de participantes e rapidez na reprodução de mensagens também podem influenciar na comunicação.

Para solucionar problemas de perda de co-texto foram desenvolvidas diversas abordagens que permitem ao usuário selecionar a mensagem a qual ele deseja responder e apresentam o diálogo na forma de árvore representando o encadeamento entre as mensagens (PIMENTEL 2002). Isso permite que os outros participantes entendam melhor a ação que a mensagem representa e o encadeamento do diálogo. Porém, Moraes (2011) destaca que a visualização do diálogo na forma de árvore dispersa o foco dos participantes dificultando a interação entre eles, além de induzir ao usuário que é necessário encadear a mensagem com alguma superior. Portanto, a perda de tempo ocasionada pela dificuldade de encadear a mensagem a ser enviada com o diálogo e a perda de atenção faz com que o batepapo perca suas características de alta interatividade e recência (FUKS et al. 2009). Devido a isso, foram desenvolvidas abordagens que apresentam o diálogo na forma cronológica e mostram o encadeamento entre as mensagens de outras maneiras. Isso faz com que o usuário perceba que o encadeamento é opcional. As ferramentas existentes que permitem o encadeamento do diálogo serão revisadas na seção 2.2.

Outro problema que ocorre em meios de comunicação textual, principalmente em diálogos que discutem assuntos complexos, é a falta de conhecimento do assunto pelos participantes.

Portanto, quando um participante não consegue acompanhar a discussão, ele procura nas mensagens anteriores a referencia em que se perdeu, busca na Web uma resposta para sua dúvida ou faz uma pergunta aos usuários sobre o assunto ou em que ponto se encontra o diálogo. Isso ocasiona a perda de tempo e prejudica o diálogo fazendo com que o usuário não consiga mais acompanhar a discussão. O entendimento equivocado de um termo ou conceito também pode causar problemas e fazer com que o usuário não consiga acompanhar o diálogo.

Assim, é desejável que os softwares de comunicação textual não sejam apenas meios para a troca de mensagens, mas também auxiliem seus usuários e proporcionem facilidades para tentar amenizar esses problemas.

#### 2.2 Análise do Diálogo

Na seção anterior os detalhes da comunicação textual e problemas que nela ocorrem foram revisados. A partir dessa discussão, surgiu a motivação para processar as mensagens

de diálogos e entender melhor como os problemas destacados acontecem e como é possível contribuir para a melhora na qualidade do diálogo. Essa etapa de processamento de mensagens tem o objetivo de analisar o diálogo gerado em ferramentas de comunicação textual visando encontrar relações entre mensagens, estruturando a comunicação. Dessa maneira, é possível perceber diversas particularidades da comunicação.

A motivação para analisar a informação gerada no diálogo surgiu da verificação dos problemas de comunicação que ocorrem nos diálogos e da necessidade de solucioná-los, além da oportunidade de extrair dados que permitam a análise e avaliação do discurso e dos participantes.

A partir da necessidade de analisar e organizar automaticamente a grande quantidade de documentos que são produzidos ao longo das discussões mediadas pelas atuais ferramentas de comunicação textual da Internet, uma metodologia chamada Comunicografia (PIMENTEL 2002) foi proposta. A pesquisa discute o emprego da metodologia em bate-papos disponíveis em ferramentas de educação à distância. Porém, ela também pode ser empregada para a análise da comunicação online em outros contextos, mudando apenas o foco das análises propostas.

A Comunicografia tem o objetivo de automatizar, totalmente ou parcialmente, a análise sobre o conjunto de textos produzidos por ferramentas de comunicação textual como: correio-eletrônico, lista de discussão, bate-papo, entre outras. Ela é desenvolvida em basicamente duas etapas: a transformação da comunicação em grafo e a análise da representação.

A representação do bate-papo no formato de grafo é chamada de Comunicografo. Esse faz uso da teoria dos grafos para representar as mensagens e as relações entre elas. Essa estrutura torna a comunicação mais compreensível, facilitando a sua análise. As mensagens são descritas a partir de um conjunto de informações geradas pela ferramenta, como o texto, nome do usuário, data e hora, ou recuperadas a partir da análise da discussão, como a referência a uma mensagem anterior, assunto, entre outros.

A representação do bate-papo no formato de grafo permite que diversos dados sejam extraídos do diálogo. Estão entre eles a identificação de interesses e conhecimentos dos participantes em determinado assunto, threads de assuntos sendo discutidos, o

31

encadeamento do diálogo, entre outras. A visualização, pelos usuários, dessa representação

permite que eles revejam o encadeamento da discussão e pontos aonde não conseguiram

acompanha-lo.

Para a elaboração do Comunicografo é necessário analisar e descobrir associações

entre as mensagens. Em (PIMENTEL 2002) são apresentados exemplos e estratégias para

identificar manualmente essas associações, porém os detalhes para a construção automática

do Comunicografo não são abordados. Segundo Pimentel, para recuperar as associações

entre as mensagens é necessário tornar explicitas as interações entre os participantes. Isso é

feito verificando padrões de coesão, coerência e participação como: referências a nomes de

usuários presentes no bate-papo, referências a palavras ou frases de mensagens anteriores,

referências a um assunto sendo discutido, entre outras.

A seguir serão discutidos trabalhos na área de análise do diálogo que propõem

métodos para identificar relações entre mensagens e podem ser utilizados como auxilio para

a definição de uma abordagem para a construção automática do Comunicografo.

Em (TRAUSAN-MATU et al. 2009), foi proposto um método para análise de bate-

papos baseado nos conceitos de polifonia linguística propostas por (BAKHTIN 1981). Nesse

trabalho foi utilizado o ConcertChat (MÜHLPFORDT et al. 2005), ferramenta de comunicação

textual em que o usuário pode escolher a qual mensagem a sua mensagem faz referência.

Assim, existe a facilidade de recuperar automaticamente referências explícitas.

Complementarmente, foram propostos métodos para recuperar o restante das referências,

implícitas. Segundo (TRAUSAN-MATU et al. 2009), os participantes só sentiam necessidade

de incluir referências explicitas quando as implícitas não eram obvias. Em (REBEDEA et al.

2008), o autor também destaca que as vezes os participantes não estão atentos ou estão

com pressa e não incluem referencias explicitas.

Um caso em que a referência foi considerada óbvia é quando aparece repetição de

palavras ou frases nas mensagens. Um exemplo do padrão é destacado a seguir:

Usuário 1: Estávamos discutindo sobre a conferência.

Usuário 2: É verdade. Vamos voltar a esse assunto.

Usuário 1: Espere um momento.

32

Usuário 2: E aonde podemos obter dados sobre essa conferência?

Outro padrão, encontrado em referências implícitas, são os pares de adjacência

(SACKS; SCHEGLOFF; JEFFERSON 1974) que são duplas de mensagens que se completam

logicamente. Alguns exemplos são os padrões pergunta-resposta e cumprimento-

cumprimento. Outro padrão encontrado são as mensagens colaborativas. Nesse caso, os

participantes constroem uma frase, como se fosse uma única mensagem. O último padrão,

encontrado no estudo, são as mensagens de convergência que fazem a ligação entre dois

assuntos diferentes discutidos nas mensagens.

No trabalho de Rebedea e colegas [2008], foi verificado que uma mensagem com

poucas palavras, concordando ou discordando, contendo no máximo uma palavra após a

retirada de Stopwords, que não faz nenhuma referencia explicita, tem uma grande

probabilidade de se referir implicitamente à última mensagem anterior a ela. Isso acontece

porque uma mensagem pequena pode ser digitada rapidamente. Um exemplo é

apresentado abaixo em uma conversa entre dois usuários:

Usuário 1: Eu acho que hoje vai chover

Usuário 2: Eu discordo

Outro método utilizado no trabalho de Rebedea foi identificar ações na discussão,

como concordar com uma mensagem. Para isso foi criada uma lista de padrões, que são

constituídos de um conjunto de palavras, aonde a referência se encontraria na mensagem.

Caso seja identificado que a mensagem corresponde a um dos padrões armazenados, a

referência é identificada e buscada em mensagens anteriores. Um exemplo seria o seguinte

diálogo:

Usuário 1: Eu já recebi minha nota do vestibular

Usuário 2: Eu não concordei com a correção do vestibular

Nesse exemplo o algoritmo encontraria o padrão "não concordei" e procuraria as

palavras "com a correção do vestibular" em mensagens anteriores. Portanto, a palavra

vestibular seria encontrada e seria verificada a existência de uma referencia implícita. Para

obter melhores resultados, o estudo citado utilizou a busca de sinônimos no Wordnet tanto

para as mensagens quanto para a lista de padrões.

O problema dos métodos desenvolvidos em (REBEDEA et al. 2008) e (TRAUSAN-MATU et al. 2009) é que as listas de padrões são definidas manualmente. Dessa forma, palavras comuns a eles podem não ser adicionadas na lista, fazendo com que algumas referências não sejam recuperadas corretamente.

Em (SHRIBERG et al. 2004), o objetivo da pesquisa é resumir automaticamente batepapos por voz, realizados por grupos de participantes, e para isso é definida a hipótese de
que encontrar concordâncias e discordâncias entre participantes contribui para a tarefa. A
abordagem, inicialmente, identifica pares de adjacência usando um algoritmo de máxima
entropia baseado em um conjunto de características léxicas, de duração, estruturais e ações
do diálogo. Os pares de adjacência identificam relações entre mensagens e são usados no
artigo para identificar a mensagem alvo de uma discordância ou concordância.

O problema, para encontrar pares de adjacência é descrito da seguinte forma: dado N candidatos, selecionar a mensagem que maximiza a distribuição de probabilidade condicional. Portanto, para isso, foi utilizada uma modelagem de máxima entropia e para treinar o modelo foram utilizados logs de diálogos manualmente anotados, identificando pares de adjacência. Portanto, para descobrir a mensagem A que é par de adjacência de uma dada mensagem B, o algoritmo busca todas as mensagens anteriores a B e utiliza o modelo de máxima entropia para identificar A.

O trabalho também destacou que informações de ações do diálogo, como perguntaresposta e oferecimento-aceitação, não contribuem significativamente como características para descobrir pares de adjacência e devido a dificuldade de obter esses dados automaticamente (STOLCKE et al. 2000) eles não foram utilizados no modelo.

Na abordagem de (GERMESIN; WILSON 2009), o objetivo também é identificar concordâncias e discordâncias em conferências de áudio e vídeo. Inicialmente, a abordagem identifica as concordâncias e discordâncias e, diferente de (SHRIBERG et al. 2004), posteriormente classifica as mensagens que são pares de adjacência para identificar para qual mensagem a concordância ou discordância está apontando. Para identificar os pares, o algoritmo também busca as mensagens anteriores e utiliza a informação de pares de adjacência para encadear as mensagens.

Outras abordagens, que visam solucionar problemas que ocorrem no diálogo, propõem ferramentas em que o usuário pode colaborar com a organização da discussão facilitando o processo de análise posterior.

As ferramentas Threaded Chat (SMITH et al. 1999) e Hiperdiálogo (PIMENTEL 2002), permitem que o usuário indique a qual mensagem ele quer responder, encadeando as mensagens na forma de árvore. Assim, elas apresentam a visualização em árvores de mensagens. A ferramenta ConcertChat (MÜHLPFORDT et al. 2005) também permite que os usuários determinem o encadeamento do diálogo, porém a visualização é apresentada na ordem cronológica e o encadeamento, visualizado ao passar o mouse na mensagem, indica a referencia entre as mensagens. A ferramenta Debatepapo (MORAES 2011) apresenta a visualização cronológica e a referência feita pela mensagem é visualizada anteriormente a ele de forma diferenciada das outras mensagens.

No AulaNet (FUKS et al. 2002) (FUKS 2000), um ambiente de aprendizagem online, as mensagens do fórum são divididas em categorias por seus autores, provendo um aspecto semântico para a relação entre as mensagens. Dessa maneira, as categorias são apresentadas ao lado do título da mensagem, possibilitando estimar como a discussão está progredindo e visualizar o objetivo de casa mensagem.

O problema dessas abordagens é que o usuário perde tempo tentando encadear as mensagens do bate-papo e escolhendo a categoria a qual a mensagem pertence. Em ferramentas de bate-papos síncronas que permitem a discussão entre grupos de usuários a velocidade da conversa necessita que o usuário esteja sempre atento ao diálogo.

Após a aplicação de métodos para análise de mensagens e construção do grafo (Comunicografo), ele é usado para facilitar a extração de estatísticas e informações interessantes sobre o diálogo. A seguir serão analisadas algumas abordagens com o objetivo de obter informações sobre a conversa após a sua organização.

Em (PIMENTEL et al. 2002), são correlacionados tópicos e autores permitindo a análise da participação e interesse de alunos. A proposta é avaliar estudantes identificando a quantidade de mensagens enviadas, o assunto e sua relevância para o diálogo. Para isso, as melhores mensagens são classificadas como aquelas que têm mais mensagens associadas, isto é, mensagens que referenciem a ela. Essas mensagens são as que incitaram mais

participantes a discutir e por isso tem mais relevância. Com a utilização do Comunicografo, também é proposta a identificação da quantidade de assuntos discutidos em paralelo, o ritmo de interação entre participantes, o tempo que demora para um participante enviar uma mensagem, quanto tempo um assunto é mantido no diálogo, entre outros.

No estudo de (TRAUSAN-MATU et al. 2009), foi proposta uma abordagem para a identificação dos assuntos mais discutidos na conversa. Para isso cada palavra encontrada nas mensagens, após a remoção de Stopwords, é representada como um conceito. Para cada conceito são identificados os sinônimos no WordNet<sup>13</sup>. Assim, os conceitos que possuem os mesmos sinônimos são agrupados em uma lista com a frequência em que cada um deles aparece durante a discussão. Dessa maneira, os sinônimos mais frequentes representam os assuntos mais discutidos.

Com a popularização das ontologias, bases de dados de palavras que representam um idioma têm sido criadas utilizando ontologias como modelo. Um exemplo é o WordNet, uma ontologia lexical que representa substantivos, verbos, adjetivos e advérbios como grupos de sinônimos. Cada um desses grupos representa um conceito distinto. Atualmente, a WordNet brasileira (DE PAIVA et al. 2012) aberta está sendo desenvolvida pela FGV<sup>14</sup>.

Em (HOLMER 2008) é proposto um método para análise e visualização da estrutura do diálogo gerado em bate-papos. A abordagem propõe uma análise manual do diálogo para identificar referencias entre as mensagens e posteriormente implementa algoritmos para analisar automaticamente o diálogo e os participantes.

Segundo Holmer (2008), identificar as relações entre mensagens é o ponto de partida para a análise estrutural da comunicação. Assim, é possível identificar discussões, dentro da comunicação, que se referem a um mesmo assunto. Fazendo a análise da estrutura gerada, dados como a contribuição de cada usuário, a coerência da comunicação e redes sociais podem ser descobertos e utilizados para avaliar a discussão.

<sup>&</sup>lt;sup>13</sup> WordNet: A lexical database for English. Disponível em <a href="http://wordnet.princeton.edu/">http://wordnet.princeton.edu/</a>>.02/08/2013

Professor da EMAp apresenta trabalho no Laboratório de IA do SRI . Disponível em:
<a href="http://emap.fgv.br/blog/2012/08/03/professor-emap-apresenta-trabalho-sri.html">http://emap.fgv.br/blog/2012/08/03/professor-emap-apresenta-trabalho-sri.html</a>>. 17/09/2012

De acordo com (PIMENTEL et al. 2002), o diálogo em um bate-papo segue a estrutura de um grafo, em que mensagens podem fazer referencias a mensagens anteriores, constituído de tópicos e sub tópicos. Em um bate-papo diferentes tópicos são discutidos ao mesmo tempo, em paralelo e alternadamente.

Como discutido anteriormente, para tornar possível a análise do diálogo com o objetivo de encadear a discussão é necessário analisar as mensagens e a relação entre elas. Portanto, a implementação do Comunicografo demandou a utilização de algoritmos de processamento de linguagem natural. Na seção seguinte os detalhes e técnicas dessa área serão revisados.

### 2.3 Processamento de Linguagem Natural

Atualmente, cresce cada vez mais a quantidade de textos gerados a cada minuto na Internet. Esses textos são criados por usuários, em linguagem natural, e são entendidos apenas por humanos, o computador não consegue processar essa linguagem. Portanto, muitos desses textos, gerados em redes sociais e blogs, ficam esquecidos. Porém, eles representam opiniões, guias e notícias extremamente úteis.

O Processamento de Linguagem Natural (PLN) é definido como um conjunto de técnicas de análise ou síntese de linguagem escrita ou falada, que podem ser tanto de hardware quanto de software. As técnicas linguísticas podem realizar análises fonéticas, relacionando a palavra com os sons produzidos, morfológicas, analisando a construção das palavras, sintáticas, relacionando as palavras entre si, e semânticas, relacionando as palavras e seus significados. Essas técnicas podem ser implementadas utilizando tecnologias conhecidas, como modelos probabilísticos, lógica, aprendizado de máquina entre outras (JURAFSKY et al. 2000).

A recuperação da informação (RI) é uma área de aplicação do PLN que visa recuperar dados em documentos, geralmente não estruturados, satisfazendo a necessidade por informação relevante, do usuário, dentro de um domínio em questão (MANNING et al. 2008).

Para facilitar o entendimento pelo computador e utilizar a informação gerada de forma

eficiente, algumas técnicas foram criadas. No Twitter é possível usar *hashtags*<sup>15</sup> para categorizar a mensagem. Dessa forma, os assuntos mais discutidos são identificados e podem ser criados gráficos e índices. No Facebook é possível mencionar pessoas, páginas e grupos para que a comunicação seja mais fácil. Essas técnicas evidenciam a necessidade de aplicação das áreas de RI e PLN para a utilização dessas informações.

Os principais passos da recuperação da informação são a operação de consulta, em que termos são especificados para representar a necessidade do usuário, a operação de indexação, em que são criadas as estruturas de dados associadas a parte textual dos documentos e a pesquisa e ordenação, em que os dados são recuperados de acordo com a consulta e é feita a ordenação de acordo com a similaridade entre a consulta e a informação.

Para recuperar documentos relevantes para uma consulta o sistema de RI adota um modelo. Muitos modelos têm sido propostos ao longo dos anos, sendo considerados como clássicos o probabilístico, o vetorial e o booleano (MANNING et al. 2008).

No modelo booleano, o índice atribuído aos documentos deve indicar qual é mais relevante do que o outro. Para isso um peso é associado a cada termo da consulta e esta é considerada como uma expressão booleana convencional formada com os conectivos lógicos AND, OR e NOT.

No modelo vetorial, documentos e consultas são representados como vetores de termos. Os documentos resultantes são representados similarmente, o vetor resultante é construído através de um cálculo de similaridade.

No modelo probabilístico, os documentos são descritos considerando pesos binários que representam a presença ou ausência de termos. O vetor gerado pelo modelo tem como base o cálculo da probabilidade de que um documento seja relevante para uma consulta.

Até esta sessão, a revisão de técnicas que utilizadas para a análise da discussão foi apresentada. Também é necessário discutir técnicas de recomendação de conteúdo que serão analisadas a partir da seção seguinte. Portanto, inicialmente, a iniciativa de dados abertos interligados será revisada para que seja possível, posteriormente, discutir a

.

Hashtags são palavras chaves antecedidas pelo símbolo #. Elas viram hiperligações dentro da rede, indexáveis pelos mecanismos de busca. Assim, ao clicar na hashtag é possível visualizar todos que também a utilizaram.

recomendação de conteúdo utilizando bases de dados em RDF como fontes.

# 2.4 A Iniciativa de Dados Abertos Interligados

Desde o seu início, a Web vem alterando a forma como o conhecimento é compartilhado. Dentre os principais formatos ainda utilizados para publicação de dados na Web, tem-se HTML, CSV, DOC e PDF. Porém, esses formatos se mostram pouco flexíveis para a associação de descritores a partes específicas de seu conteúdo, assim como não permitem diretamente uma descrição da estrutura e da semântica dos dados. A abordagem de dados abertos interligados se apoia na proposta de boas práticas para expor, compartilhar e conectar dados, informações e conhecimento na Web Semântica utilizando URIs e RDF.

O RDF é utilizado nessa abordagem para representar o significado das informações, representado-as no formato de triplas. Essas triplas são constituídas de um sujeito, um predicado e um objetivo. O sujeito e o objeto podem ser representados por identificadores únicos chamados de URIs. Existem casos em que o objeto não é um recurso e é representado por um literal. Os predicados representam a relação entre o sujeito e o objeto.

O memo<sup>16</sup> de Tim Berners Lee publicado no W3C, em 2007, estabelece quatro regras para a publicação de dados abertos na Web de modo que eles possam ser ligados a outros dados. São elas:

- 1. Usar URIs como nomes de recursos
- 2. Usar URIs HTTP de modo com que pessoas possam buscar esses nomes
- 3. Quando alguém buscar uma URI, mostre informação útil, utilizando os padrões(RDF\*, SPARQL)
- 4. Incluir ligações para outras URIs para que seja possível descobrir outros recursos

A iniciativa de dados abertos interligados, apoiada pelo W3C e seguida por muitas instituições mundialmente, vem identificando bases de dados abertas existentes convertendo-as para RDF, de acordo com os princípios de LOD e publicando-as na Web. De

\_

Berners-Lee, T. (2006). Linked Data - Design Issues Disponível em: <a href="http://www.w3.org/DesignIssues/LinkedData.html">http://www.w3.org/DesignIssues/LinkedData.html</a>. 14/04/2012

acordo com as estatísticas de setembro de 2011, coletadas pelo projeto LATC<sup>17</sup>, existiam até aquela data 295 fontes, contendo mais de 31 bilhões de triplas RDF, no total, interligadas por mais de 503 milhões de ligações RDF.

Para que seja possível consumir as bases de dados disponíveis de acordo com os princípios de LOD, a linguagem SPARQL foi criada. Ela é o padrão definido para a consulta de informações presentes em bases de triplas em RDF. Para que seja possível consumir essa informação remotamente, sem a necessidade de obter a base de dados, geralmente esses dados são acessíveis através de um serviço Web (HARTIG et al. 2009). Esse serviço tem o nome de SPARQL Endpoint e permite a execução de consultas SPARQL. Assim, é possível acessar o serviço através de uma página Web ou de APIs disponíveis para diversas linguagens de programação.

A seguir é mostrado um exemplo do formalismo RDF e uma consulta SPARQL realizada para recuperar esses dados.

<rdf:Description rdf:about="http://pt.dbpedia.org/resource/Dilma\_Rousseff">

<dbpedia-prop:nome> Dilma Rousseff</dbpedia-prop:nome>

<dbpedia-owl:religion rdf:resource="http://pt.dbpedia.org/resource/Catolicismo" />
</rdf:Description>

O RDF acima, encontrado na DBpedia, descreve a presidente Dilma Rousseff, indicando seu nome e sua religião. A DBpedia em português disponibiliza um SPARQL Endpoint no endereço http://pt.dbpedia.org/sparql. Portanto, para descobrir a religião de Dilma Rousseff seria executado o SPARQL:

```
SELECT DISTINCT ?religiao WHERE {
    <a href="http://pt.dbpedia.org/resource/Dilma_Rousseff">http://pt.dbpedia.org/resource/Dilma_Rousseff</a>> dbpedia-owl:religion ?religiao .
}
```

A evolução da Web Semântica e da iniciativa de LOD não aconteceu apenas através do aumento do número de bases publicadas, mas também pela criação de diversas

-

Bizer, Jentzsch, Cyganiak. State of the LOD Cloud Disponível em <a href="http://www4.wiwiss.fu-berlin.de/lodcloud/state/">http://www4.wiwiss.fu-berlin.de/lodcloud/state/</a>>.14/04/2012

ferramentas para auxiliar a publicação, ligação e utilização desses dados. Atualmente, existem diversos sistemas de bancos de dados para o armazenamento de recursos em RDF e ferramentas para transformar dados de diversos formatos para RDF, para descobrir relações entre bases de dados distintas e fazer ligações, e aplicações para visualização desses dados.

As aplicações que exploram a WoD podem ser divididas em três categorias: navegadores, ferramentas de busca e aplicações específicas de domínio. Entre as aplicações específicas de domínio, pode-se citar a aplicação El Viajero (GARIJO; VILLAZÓN-TERRAZAS; CORCHO 2011) que agrega conteúdo em RDF de diversos jornais, plataformas digitais, recomendações de usuários, notícias de blogs e imagens. Com isso, os usuários podem explorar, reusar e organizar essas informações. Os dados de localização, presentes nas notícias, permitem que elas sejam visualizadas em um mapa, e a utilização de informações inseridas pelos usuários permite a visualização de fotos, vídeos e comentários de uma viagem. Outro exemplo interessante é a BBC (KOBILAROV et al. 2009), que utiliza dados da Dbpedia para enriquecer suas informações, e usa RDF para publicar dados de diversos domínios e interligá-los.

# 2.5 Sistemas de Recomendação

Como será detalhado no capítulo 3, essa pesquisa trabalha com recomendação de conteúdo e recuperação da informação de dados da Web de Dados em ambientes de comunicação textual online. Portanto, a partir da apresentação da iniciativa de dados abertos interligados, realizada na seção anterior, neste tópico são revisadas abordagens para recomendação de conteúdo da Web de Dados e abordagens para recomendação de conteúdo em ferramentas de comunicação textual.

Os sistemas de recomendação fornecem informação personalizada para usuários, de acordo com seus interesses e o contexto em que estão inseridos. Geralmente, esses sistemas possuem como entrada conceitos importantes para o usuário e retornam uma lista de conceitos ordenados por similaridade que podem ser interessantes para ele. No cenário atual eles têm grande importância devido ao aumento da quantidade de informação disponível na Web e em sistemas privados como as plataformas de e-commerce. Dessa maneira, os sistemas de recomendação tem o objetivo de selecionar conteúdo de acordo com as preferências do usuário.

Esses sistemas utilizam informação de diversas fontes, dependendo do conteúdo que será recomendado. Segundo Passant, Heitmann e Hayes (2009), a maioria dos sistemas de recomendação utiliza apenas informação privada ou informação pública gerada colaborativamente, que, geralmente, só contém dados sobre um domínio específico. É o caso da Amazon que utiliza dados sobre os usuários e produtos cadastrados no site para fazer recomendações.

Os sistemas de recomendação podem ser classificados basicamente em três categorias: baseados em conteúdo, filtragem colaborativa e sistemas híbridos (DOMAVICIUS; TUZHILIN 2005). Os sistemas baseados em conteúdo utilizam técnicas de RI para recuperar itens a partir das preferencias do usuário. Assim, o sistema classifica os itens de acordo com a similaridade destes com os classificados como importantes para o usuário anteriormente. A filtragem colaborativa consiste em recomendar itens baseados na similaridade entre usuários. Basicamente, são selecionados usuários que escolheram, ou tem preferencia, por itens similares ao usuário alvo. Assim, os itens recomendados são os que foram bem classificados por esse grupo de usuários. Os sistemas híbridos consistem na associação das técnicas apresentadas anteriormente visando unir as vantagens e reduzir os problemas encontrados em cada uma delas. Portanto, atualmente é muito comum que sistemas de recomendação utilizem um grupo de conceitos para a busca de itens.

Considerando o cenário atual da Web Semântica é possível construir sistemas de recomendação, baseados em diversos domínios, utilizando fontes de informação da WoD. Isso é possível devido à interoperabilidade entre essas bases de dados e a reutilização de ontologias para descrição dos dados tornando possível o desenvolvimento de uma única consulta que possa ser executada em diversas bases. De acordo com Passant, Heitmann e Hayes (2009), a construção de sistemas baseados em dados representados em LOD e na linguagem de busca SPARQL permite a utilização de apenas um algoritmo de recomendação por sobre diversas fontes de dados.

### 2.5.1 Recomendação em Ferramentas de Comunicação Textual

Atualmente, já existem algumas abordagens para recomendação de conteúdo em ferramentas de comunicação textual. Loh e colegas (2010) construíram um sistema de recomendação que analisa o texto sendo discutido em um bate-papo e recomenda

informação complementar baseada no contexto da discussão e nos perfis dos usuários. O sistema utiliza uma ontologia para identificar conceitos nas mensagens, empregando uma lista de palavras-chaves para cada conceito da ontologia, e classificar uma base de documentos que serão recomendados dentro desses conceitos. Dessa maneira, com base no diálogo, pessoas e assuntos são relacionados permitindo que documentos de uma base préexistente sejam recomendados durante o bate-papo.

O GaChat (HORIGUCHI et al. 2009) é uma ferramenta de discussão online, para a língua japonesa, que recomenda imagens e artigos, recuperados da Web, relacionados a substantivos utilizados na discussão. Assim, o objetivo dessa abordagem é solucionar dúvidas do usuário fazendo com que ele adquira conhecimento na própria ferramenta.

Chandra, Cambria e Pradeep (2011) desenvolveram uma ferramenta de bate-papo que mostra condições meteorológicas, como uma metáfora, para o estado de humor do texto. Para isso as mensagens são analisadas utilizando técnicas de análise de sentimento, a partir desse resultado o plano de fundo do bate-papo muda mostrando a condição meteorológica correspondente.

Nesses trabalhos são recomendadas apenas informações de uma única base e nenhuma delas utiliza dados da WoD. Portanto, existe a necessidade de uma abordagem mais abrangente que traga benefícios para a comunicação em uma diversidade de domínios.

A identificação de elementos relevantes em um texto constitui ponto fundamental das estratégias de recomendação, para que sirvam como referência na busca por informações complementares. Portanto, no caso da comunicação textual é extremamente importante identificar dados como: o assunto da conversa, pontos de dúvidas, usuários que não conseguem acompanhar a discussão, entre outros.

#### 2.5.2 Recomendação de Dados Abertos Interligados

Segundo descrito acima, a partir de possibilidade de utilizar dados disponíveis na WoD para recomendação de conteúdo foram implementados algoritmos para tal, que utilizam técnicas de recomendação e recuperação da informação, e serão descritos a seguir.

Passant, Heitmann, Hayes (2009) desenvolveram o algoritmo LDSD (Linked Data Semantic Distance) para a recomendação de conteúdo de bases de dados em RDF. O

algoritmo mede a distância entre entidades, dependendo do número de relações diretas entre elas. Uma segunda versão do algoritmo, além de verificar o número de relações diretas entre entidades, também verifica o número de relações indiretas (PASSANT 2010), aquelas aonde além de se verificar se o recurso A está relacionado com B, também é verificado se A e B tem, os dois, relação com C. Como exemplo, no domínio de músicos, pode-se tanto verificar se dois cantores já compuseram, em conjunto, uma canção, quanto se eles participaram de um mesmo festival.

Mirizzi e colegas (2012) adaptaram o algoritmo Vector Space Model (VSM) (SALTON; WONG; YANG 1975) para a recomendação de dados em RDF. No VSM, pesos não binários são atribuídos aos termos para calcular a similaridade. Dessa maneira, as relações são exploradas semanticamente, pois diferentemente do algoritmo LDSD, cada relação tem uma importância distinta. A aplicação MORE (More than movie recomendation) foi desenvolvida pelos mesmos autores, utilizando o algoritmo VSM. A ferramenta utiliza dados da DBpedia, para recomendar filmes a usuários do Facebook.

De acordo com essa visão, a aplicação dbrec (PASSANT 2010), um sistema de recomendação de música que utiliza dados da DBpedia, oferece recomendações de mais de 39.000 bandas e cantores. Nela, o algoritmo de recomendação LDSD é utilizado para calcular a similaridade entre dois recursos. Uma vantagem do RDF é que, nesse caso, foi possível criar uma ontologia que permite a publicação dos dados de similaridade e da descrição de como eles foram calculados. Ainda nessa linha, Shabir e Clarke (2009) desenvolveram um sistema de recomendação para aprendizado, chamado Talis Aspire. O sistema gera triplas RDF a partir das listas de recursos criadas pelos usuários para determinado curso. Isto permite que itens presentes em uma lista sejam ligados a itens correspondentes em listas de outras instituições, construindo uma rede de dados acadêmicos.

# 3 AidChat – Sistema de Recomendação para Bate-papos

A abordagem proposta e implementada nesse trabalho tem o objetivo de analisar mensagens enviadas em uma ferramenta de comunicação textual online, extrair dados do diálogo e apresentar informações interessantes que possam contribuir com o conhecimento do usuário sobre o assunto e consequentemente com a qualidade da comunicação que será medida a partir da avaliação da relevância das informações pelos participantes. Embora tomando como base o cenário de suporte à discussão da legislação brasileira, com pequenas modificações, também é possível aplicá-la em outros cenários.

Esse capítulo está estruturado da seguinte forma: a primeira seção define os requisitos para a definição e implementação da abordagem, a segunda descreve conceitualmente a abordagem e a terceira apresenta a arquitetura e detalhes da implementação.

# 3.1 Definição de Requisitos

A partir dos trabalhos relacionados revisados nessa pesquisa é possível definir os requisitos para a implementação da abordagem proposta. De acordo com os problemas encontrados em bate-papos com grupos de participantes, principalmente no cenário estudado, e considerando dificuldades apontadas por trabalhos relacionados revisados foi possível perceber que um problema muito comum, nos diálogos em contextos complexos com grupos de participantes, são as dúvidas e falta de conhecimento do assunto pelos participantes. Tais problemas ainda não foram muito abordados em trabalhos relacionados e carecem de análise e solução.

Portanto, visando enriquecer a experiência de comunicação online textual, acreditase que recomendações de conteúdo, relacionadas ao contexto sendo discutido, podem reduzir as dúvidas dos participantes. Dessa maneira, o diálogo se torna mais rico e construtivo pois os usuários passam a obter conhecimento na própria ferramenta adquirindo capacidade de argumentação e entendimento do diálogo.

O primeiro requisito percebido é entender o encadeamento do diálogo, identificando referências e relações entre as mensagens. Além disso, a partir da análise da estrutura da

discussão é necessário apoiar o entendimento da sequencia de discussão sobre um mesmo tema, identificando tópicos e subtópicos. Essa análise inicial torna o bate-papo mais compreensível e permite que sejam identificados detalhes como a quantidade de tópicos sendo discutidos ao mesmo tempo, os participantes que interagiram mais em cada um deles e quais os que foram mais discutidos.

Também é necessário identificar dúvidas ocorridas e os momentos e frequência em que elas ocorrem. Além disso, é interessante verificar a atitude dos participantes ao perceber um conceito que não entederam ou perceber que não estão acompanhando o encadeamento do diálogo.

Visando a recomendação posterior, além de encontrar dúvidas, é necessário identificar termos relevantes sobre o contexto sendo discutido. A partir disso, é possível analisar as informações que seriam interessantes como contribuição à discussão.

Assim, também é importante que as informações recomendadas sejam sucintas, permitindo que os participantes possam lê-las rapidamente, além de serem suficientes para que o participante consiga entender o conceito e continuar acompanhando a discussão. Dessa maneira, é importante que as recomendações também sejam apresentadas em um curto espaço de tempo, antes que o assunto do diálogo mude e os participantes não possam mais argumentar sobre ele.

Outra questão interessante é identificar como dados relevantes podem ser extraídos do diálogo, permitindo que essa informação possa ser reutilizada. Assim, o conteúdo recomendado pode ser utilizado em diálogos realizados posteriormente, no mesmo contexto. Além disso, também podem ser extraídas informações sobre os participantes permitindo que elas sejam utilizadas, por exemplo, para recomendar artigos no contexto de assuntos interessantes para ele.

A seguir, ao longo da análise da abordagem desenvolvida e da sua arquitetura, os detalhes para a satisfação dos requisitos apresentados serão abordados.

### 3.2 Abordagem Proposta

A abordagem proposta, definida a partir dos requisitos, se apoia na metodologia de Comunicografia que define técnicas para a análise do diálogo e a extração de informações sobre ele. Ela facilita o entendimento da sequencia de discussão a partir da sua representação no formato de grafo e propõe avaliações que podem ser feitas a partir dele, extraindo informações com relação ao diálogo e aos participantes.

Posteriormente, a partir da análise do Comunicografo são encontradas mensagens em que dúvidas ocorreram, permitindo que elas sejam usadas posteriormente para encontrar conteúdo relevante para recomendação.

Portanto, a partir do entendimento do diálogo e a identificação dos momentos em que dúvidas ocorreram, informações relacionadas são recomendadas. Estas visam agregar conhecimento ao diálogo permitindo que dúvidas sejam solucionadas, aumentando a qualidade e resultados deste.

Nessa abordagem, dados da WoD são utilizados como fonte de recomendação. Estes permitem que sejam utilizadas bases de dados já existentes e, muitas das vezes, interligadas com outras bases, possibilitando uma grande abragencia de conteúdos. Além disso, a estrutura definida no método de LOD permite a preparação de dados extraídos do diálogo para a sua publicação posterior na Web. Isso possibilita que as informações recomendadas e dados referentes aos participantes agregem a WoD, permitindo a sua utilização em batepapos posteriores e até em outras aplicações.

A seção seguinte apresenta a arquitetura de apoio à abordagem e os detalhes para a sua implementação.

# 3.3 Arquitetura de Apoio à Abordagem

Tendo como objetivo a evolução dos meios de comunicação textual online, a arquitetura da abordagem proposta foi dividida em três módulos principais. São eles: préprocessamento de mensagens, análise do diálogo e recomendação de conteúdo (figura 9). Para enriquecer a experiência de comunicação, o AidChat executa os módulos após cada mensagem enviada ao longo do diálogo.

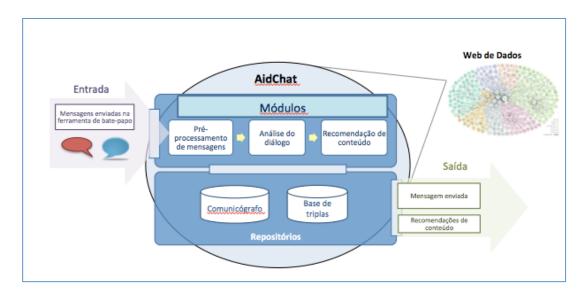


Figura 9: Arquitetura do AidChat

Portanto, a abordagem proposta tem como entrada a mensagem enviada pelo participante, que é processada antes de sua exibição na tela. Inicialmente, é executado o módulo de pré-processamento para tratar a mensagem enviada com o objetivo de facilitar a análise posterior. O módulo de análise do diálogo tenta relacionar as mensagens, possibilitando a geração de um grafo, chamado de Comunicografo e a percepção das threads de mensagens relacionadas que representam tópicos discutidos e pontos de dúvida que são armazenados em conjunto. O grafo gerado é armazenado em arquivos XML, portanto, o AidChat armazena a mensagem enviada e a referência que ela faz a uma mensagem anterior, se houver. Para identificar essa referência são aplicados algoritmos, descritos na seção 3.3.2, que levam em consideração a mensagem enviada e as uma quantidade de mensagens anteriores, definida na seção 3.3.2.3, já armazenadas no arquivo XML. O módulo de recomendação de conteúdo tem o objetivo de utilizar os dados gerados na análise, isto é, o Comunicografo, para recomendar conteúdo interessante, recuperado de bases de triplas da WoD, complementando, assim, o diálogo. Finalmente, após a aplicação dos módulos desenvolvidos, o AidChat tem como resultado a mensagem submetida em conjunto com informações, no formato de um mapa chave/valor, provenientes da Web de Dados, que contribuem construtivamente com o diálogo.

#### 3.3.1 Módulo de Pré-processamento de Mensagens

A comunicação online, em todo o mundo, tem proporcionado uma nova linguagem que foge da norma culta. Por ser um ambiente em que a escrita é muito utilizada, requer

habilidade e velocidade para atender ao estilo de vida no mundo globalizado. Portanto, o usuário, ao interagir com seus interlocutores, nos meios de comunicação textuais, necessita fazer uso de uma escrita com características que a aproximam da oralidade.

Devido às adequações da linguagem que a Web necessita, nos bate-papos informais, os usuários usam reduções, abreviações, símbolos e gírias. Tudo isso para facilitar a rapidez nas comunicações, permitindo que façam diversas tarefas ao mesmo tempo. No dia a dia é comum que as pessoas interajam com vários usuários ao mesmo tempo em bate-papos, utilizem uma rede social e façam pesquisas em *websites* em um curto espaço de tempo. Porém, quando o contexto da discussão é mais sério, a quantidade de abreviações e gírias utilizadas geralmente é menor, tornando a análise e processamento do texto mais simples.

Devido a isso, para facilitar a análise posterior é necessário que seja feito o processamento inicial do texto utilizando as técnicas de processamento de linguagem natural.

Portanto, o módulo de pré-processamento de mensagens tem o objetivo de analisar as mensagens, corrigir problemas de escrita e remover palavras irrelevantes para facilitar a análise posterior da discussão. Para esses fins serão utilizadas técnicas de processamento de linguagem natural.

#### 3.3.1.1 Detalhamento

O pré-processamento é a primeira etapa a ser realizada para que os dados possam ser analisados. O texto encontrado em mensagens trocadas a partir de softwares para a comunicação textual online geralmente apresenta erros ortográficos devido a velocidade de digitação exigida, que aumenta ainda mais quando a quantidade de participantes é muito grande.

Outro problema existente nas mensagens enviadas nos bate-papos é a quantidade de abreviações, símbolos e gírias utilizadas. Geralmente, os participantes utilizam símbolos que representam emoções, abreviações para que sea possível digitar mais rapidamente e gírias comuns ao meio em que estão discutindo.

Portanto, é necessário solucionar as inadequações no uso da linguagem que comprometem a compreensão do texto, a análise do diálogo e a recomendação de

conteúdo. Para isso, é feita a correção ortográfica das mensagens enviadas na ferramenta.

Além disso, para que seja possível analisar a mensagem também é necessário remover termos que são irrelevantes para a análise. As palavras com alta frequência e baixa relevância no texto atrapalham o processo de análise e recuperação da informação. Um exemplo é a verificação dos tópicos mais discutidos que utiliza a recuperação das palavras que tem maior frequência no texto.

Portanto, outra necessidade no pré-processamento é a remoção de *Stopwords*. *Stopwords*, ou palavras de parada, são palavras que podem ser consideradas irrelevantes para a consulta ao conjunto de resultados a ser exibido. As *Stopwords* geralmente são artigos, preposições e conjunções. Alguns exemplos dessas palavras são: as, e, os, de, para, com, sem, foi. Também é possível adicionar a essa lista outras palavras que sejam irrelevantes para o cenário sendo discutido.

Outra necessidade verificada na análise do diálogo foi reduzir as variações dos radicais das palavras, permitindo que palavras com radicais diferentes fossem consideradas iguais quando comparadas.

A radicalização ou *stemming* é uma técnica que visa reduzir variações de um radical para converter as diferentes variações de uma palavra a uma forma básica denominada raiz ou stem com o objetivo de buscar palavras correlatas. Portanto, a radicalização reduz as palavras ao seu radical por meio da retirada dos afixos (sufixos e prefixos).

Considerando os problemas descritos, no módulo de pré-processamento foram implementados os seguintes algoritmos: correção ortográfica, remoção de Stopwords e radicalização. Os detalhes para a implementação destes são discutidos na seção seguinte.

## 3.3.1.2 Implementação

Segundo a seção anterior, o primeiro passo desse módulo é a aplicação de um corretor ortográfico. Atualmente, existem APIs em Java e dicionários de sinônimos que podem ser utilizados para identificar erros ortográficos. Dessa maneira, é possível utilizar dicionários genéricos e dicionários que representem o cenário sendo discutido para que o texto seja corrigido corretamente.

Portanto, para que a os erros ortográficos e erros de digitação sejam corrigidos foi utilizada uma API para a linguagem Java chama Jazzy<sup>18</sup>. Ela é uma API de código aberto que utiliza um dicionário para a verificação de palavras. Como dicionário foi utilizado o br.ispell<sup>19</sup>, um dicionário para português, para o vocabulário utilizado no Brasil.

No segundo passo é realizada a remoção de Stopwords. A obtenção da lista de Stopwords pode ser manual dependo da necessidade e contexto do sistema. Porém, existem diversas listas e APIs que reúnem as palavras mais comuns e também podem ser utilizadas.

O Lucene<sup>20</sup> é uma biblioteca do Apache, para indexação e buscas textuais, de código aberto e desenvolvida em Java. Ela é constituída por duas etapas principais: a indexação e a análise. A indexação é o processo de conversão do texto em um formato eficiente para a pesquisa por palavras-chave. Na análise várias operações são executadas como: extração das palavras, remoção de palavras comuns, remoção da pontuação, redução de palavras para o formato de raiz, alteração das palavras para minúsculas, entre outras.

Portanto, no Lucene a extração do texto é feita a partir de um Analyser, classe que contém as regras para extrair conteúdo. No entanto, existem diversas implementações dessa classe que realizam essa função. A classe utilizada na abordagem é a BrazilianAnalyzer, desenvolvida para o português do Brasil, que contém uma lista de Stopwords em português. A lista contida nessa classe inclui artigos, pronomes, advérbios, preposições, conjunções, consoantes e vogais. Após a remoção de Stopwords é criado um índice, com o texto extraído, que pode ser utilizado posteriormente para consulta.

A próxima etapa é a implementação de um algoritmo de radicalização. Tradicionalmente utiliza-se o radicalizador de Porter (PORTER 1980), desenvolvido especificamente para a língua inglesa. Porém, esse algoritmo baseado em regras remove apenas os sufixos e não funciona muito bem para língua portuguesa.

A implementação do algoritmo Portuguese Stemmer, proposto por Viviane Orengo e Christian Huyck em (ORENGO; HUYCK 2001) é composta por oito passos em que cada passo

<sup>&</sup>lt;sup>18</sup> Jazzy The Java Open Source Spell Checker. Disponível em: <a href="http://jazzy.sourceforge.net/">http://jazzy.sourceforge.net/</a>>. 03/02/2013

<sup>&</sup>lt;sup>19</sup> Dicionário br.ispell. Disponível em: <a href="http://www.ime.usp.br/~ueda/br.ispell/">http://www.ime.usp.br/~ueda/br.ispell/</a>>. 04/11/2013

<sup>&</sup>lt;sup>20</sup> Lucene. Disponível em: <a href="http://lucene.apache.org/core/">http://lucene.apache.org/core/</a>>. 04/11/2013

tem seu conjunto de regras. As regras dentro de cada passo são examinadas em sequência, somente uma delas é aplicada e o sufixo mais longo possível é sempre removido primeiro. As etapas realizadas nesse algoritmo são: redução do plural, redução do feminino, redução adverbial, redução do aumentativo/diminutivo, redução nominal, redução verbal, redução de vogais e redução de acentos. O algoritmo também utiliza listas de exceções que ocorrem na língua portuguesa. Por exemplo, -ão é um sufixo comumente usado para indicar aumentativo, contudo nem todas as palavras terminadas em -ão denotam aumentativo.

Existe também uma versão do algoritmo de Porter para a língua portuguesa<sup>21</sup>. O algoritmo também está baseado em regras. São elas:

- remoção dos sufixos;
- remoção dos sufixos verbais, se o primeiro passo não realizou nenhuma alteração;
- remoção do sufixo i, se precedido de c;
- remoção dos sufixos residuais os, a, i, o, á, í, ó;
- remoção dos sufixos e, é, ê e tratamento da cedilha.

Portanto, o algoritmo Portuguese Stemmer apresenta uma melhora em relação ao de Porter adaptado ao português por tratar exceções que podem ocorrer na língua. De acordo com Viera et al (2006), o algoritmo de Porter reduz o tamanho do vocabulário em 44%, enquanto o algoritmo de Orengo reduz o tamanho em até 51% e o último gera menos erros.

De acordo com a análise realizada, o algoritmo utilizado para radicalização foi o Portuguese Stemmer. Foi utilizada a implementação da API do Weka<sup>22</sup>. O Weka é um software livre escrito em Java que disponibiliza um conjunto de algoritmos de aprendizado de má quina para mineração de dados. Os algoritmos podem ser utilizados a partir do software ou diretamente da API em Java. As bibliotecas utilizadas para radicalização estão localizadas em weka.core.stemmers e a biblioteca utilizada foi a PTStemmer.

Após a aplicação dos algoritmos de limpeza e pré-processamento, a mensagem enviada está preparada para ser analisada.

<sup>22</sup> Weka 3: Data Mining Software in Java. Disponível em: <a href="http://www.cs.waikato.ac.nz/ml/weka/>.10/03/2013">http://www.cs.waikato.ac.nz/ml/weka/>.10/03/2013</a>

<sup>&</sup>lt;sup>21</sup> Porter, M. (2005). Snowball. Disponível em: <a href="http://snowball.tartarus.org/index.php">http://snowball.tartarus.org/index.php</a>>. 10/11/2012

## 3.3.2 Módulo de Análise do Diálogo

Após a aplicação dos algoritmos de pré-processamento, a mensagem submetida, pré-processada, e as vinte mensagens anteriores, previamente armazenadas no Comunicografo, são processadas pelo módulo de análise (figura 10). Nesse módulo, inicialmente, são aplicados algoritmos para verificar referências implícitas, pares de adjacência, entre a mensagem atual e uma das vinte anteriores. Para isso, um algoritmo estatístico foi implementado de acordo com as características relevantes para o encadeamento do diálogo. Dessa maneira é possível estruturar o diálogo na forma de grafo e armazená-lo no Comunicografo, permitindo sua análise posterior.

Após a estruturação das mensagens, uma lista de padrões, que contém palavras que representam falhas de comunicação, é utilizada para identificar quando os usuários apresentaram dúvidas na mensagem submetida. Essa lista foi criada a partir da análise de logs de bate-papos coletados e descritos na seção 3.3.2.4. O algoritmo desenvolvido marca a mensagem no Comunicografo, indicando que ela contém uma dúvida, para que posteriormente seja possível recuperar informações relevantes que ajudem o usuário a entendê-la.

Posteriormente, o módulo de recomendação, descrito na seção 3.3.3, utiliza o Comunicografo armazenado e a mensagem enviada para buscar informação relevante para os participantes em bases de dados pré-definidas.

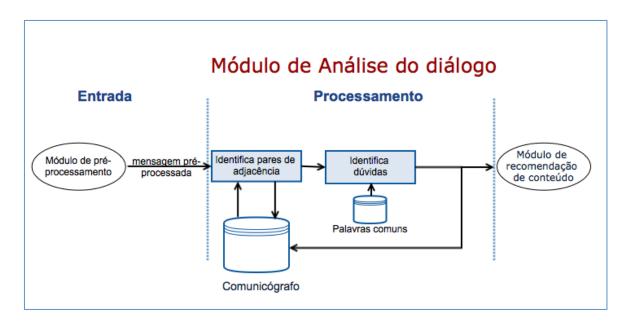


Figura 10: Diagrama do módulo de análise do diálogo

### 3.3.2.1 Detalhamento do Comunicografo

A implementação de um algoritmo para a construção automática do Comunicografo demandou que fossem implementados algoritmos para identificar referencias entre mensagens. Para isso definiu-se um modelo bastante simples para representar a discussão (figura 11). Ele é utilizado para organizar e armazenar as mensagens enviadas durante a conversa.

Nesse modelo o diálogo é representado utilizando as classes discussão, mensagem e usuário. A classe discussão representa o diálogo com um todo e faz referência para a lista de mensagens enviadas. A classe mensagem tem associações com o usuário que a enviou, a discussão em que ela foi gerada e a mensagem a qual ela faz referência. Essa classe tem como atributos o texto enviado, a data e o tipo, que é a ação que a mensagem representa. São exemplos dos tipos de ações que a mensagem pode exercer pergunta, resposta, discordância, proposta, aceitação e etc. A classe usuário possui como atributo o seu apelido no bate-papo.

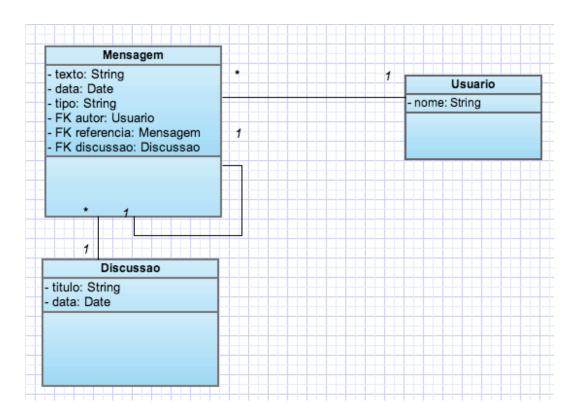


Figura 11: Modelo de dados para apoio à representação da discussão

Portanto, a partir de uma lista com as vinte últimas mensagens enviadas, recuperadas a partir do Comunicografo previamente armazenado, a ferramenta aplica os algoritmos de forma ordenada e ao identificar uma associação, salva a mensagem de referencia no objeto mensagem atualizando o Comunicógrado armazenado.

O modelo definido permite que cada mensagem faça referência a apenas uma outra mensagem. Isso é feito como meio de simplificação e a associação da mensagem com a mais próxima a ela, em que a referência é encontrada, é representada. Essa simplificação é necessária, pois a ocorrência de duas associações entre as mensagens pode acarretar na formação de ciclos (PIMENTEL 2002).

Dessa forma, quando uma mensagem é enviada, o algoritmo percorre as mensagens a partir dela até as vinte anteriores, procurando a mensagem a qual a última mensagem enviada faz referência. No momento em que uma associação é encontrada, o algoritmo espera que a próxima mensagem seja enviada, pois cada mensagem só pode fazer uma referência. Segundo Pimentel (2002), esses problemas não ocorrem com muita frequência e simplificação pode ser feita para facilitar a análise posterior.

Existem alguns casos em que a mensagem se refere a mais de uma mensagem

anterior. Isso acontece, por exemplo, quando o usuário responde duas perguntas em uma mesma mensagem. Portanto, o impacto dessa simplificação é que esse tipo de mensagem só será associado com a mensagem mais significativa, encontrada pelo algoritmo, a qual ela faz referência.

Portanto, para identificar referencias entre as mensagens, um modelo estatístico que verifica diversos padrões entre mensagens foi desenvolvido. O modelo permite que se recupere as associações que ocorrem no diálogo gerando um grafo que o represente. Essas associações entre as mensagens são chamadas de pares de adjacência.

Pares de adjacência são constituídos de duas mensagens que referenciam uma a outra e são enviadas por usuários diferentes. De acordo com (SACKS; SCHEGLOFF; JEFFERSON 1974), pares de adjacência são considerados unidades fundamentais na organização da conversa. A sua identificação é muito importante, pois permite identificar referências entre mensagens como as ações: pergunta e resposta, cumprimento e cumprimento, concordâncias, discordâncias, entre outras.

Nos diálogos por voz, de acordo com as regras típicas da oralidade praticadas implicitamente, normalmente, os pares de adjacência são adjacentes, isto é, eles ocorrem seguidamente. Porém, na comunicação textual, principalmente quando há mais de dois de participantes, isso não é verdade. Portanto, para encontrar o par de adjacência de uma mensagem é necessário percorrer as mensagens anteriores enviadas por outros usuários procurando uma referencia entre elas.

O número de mensagens anteriores a ser percorrido foi determinado de acordo com os bate-papos anotados nesse trabalho posteriormente, na seção 3.3.2.4. Portanto, as distâncias entre pares de adjacência foram medidas em todos os logs anotados. A maior distância entre pares de adjacência foi de quinze mensagens. Porém, como os diálogos anotados não contemplam todas as possibilidades, pode existir um bate-papo em que a distância seja maior que essa. Portanto, foi estabelecido um valor um pouco maior e o algoritmo percorre sempre as vinte mensagens anteriores buscando uma referencia.

Foram desenvolvidas duas versões de implementações desse algoritmo, pois a primeira não apresentou resultados satisfatórios. Os detalhes dessas implementações serão discutidos nas seções 3.3.2.2 e 3.3.2.3.

## 3.3.2.2 Implementação da Primeira Versão

Essa implementação foi desenvolvida de acordo com análise da revisão de literatura apresentada na seção 2.2. O algoritmo desenvolvido nessa seção percorre as vinte mensagens anteriores a cada mensagem enviada procurando pela repetição de alguma palavra. Devido à utilização de técnicas de pré-processamento de mensagens, ele encontra palavras com radicais iguais.

Para exemplificar o caso de repetição, o trecho de um diálogo realizado no bate-papo do site e-Democracia, que discute o assunto médicos do exterior no Brasil, durante uma assembleia legislativa é destacado na figura 12.

```
(12/06/2013 - 10:44:02) - rodrigosetubal fala : Não estou tendo acesso ao video. Vcs estão? (12/06/2013 - 10:44:10) - Silvestre SOUSA Sousa fala : Bom dia senhores! (12/06/2013 - 10:44:52) - Comissão de Constituição e Justiça e de Cidadania fala : O vídeo somente pode ser acessado pelo Internet Explorer.
```

Figura 12: Exemplo de repetição de palavras em um bate-papo

O bate-papo apresentado, um participante pergunta sobre o acesso ao vídeo da assembleia. Logo abaixo, a comissão faz referência à palavra vídeo como resposta a pergunta feita pelo participante. Porém, o algoritmo é falho nos casos em que o usuário se refere a uma mensagem anterior utilizando sinônimos.

Dessa maneira, para aumentar a eficácia dessa implementação poderiam ser utilizados sinônimos de cada palavra para buscar referências. Essa técnica foi utilizada em (TRAUSAN-MATU et al. 2009) para identificar os assuntos discutidos no diálogo. Em batepapos, muitas vezes um usuário se refere a uma palavra presente em outra mensagem utilizando um sinônimo. Para recuperar sinônimos de uma palavra é possível utilizar um dicionário como o Wordnet. Porém, só existe uma versão do Wordnet disponível para utilização na língua portuguesa, o OpenWordnet-PT (DE PAIVA et al. 2012). Entretanto, esta base de termos ainda está em fase de desenvolvimento e ainda não produz resultados satisfatórios quando utilizada para a busca de sinônimos.

De acordo com os testes realizados, essa primeira versão do algoritmo também é falha em outras situações. Em bate-papos online, como destacado em (PIMENTEL et al. 2002), é comum que o usuário faça referência a uma mensagem citando o nome do participante que a enviou. A figura 13 apresenta um exemplo destacado de um bate-papo

que ocorreu no site e-Democracia simultâneo a uma audiência pública para discutir a reformulação do ensino médio.

```
(28/11/2012 - 12:50:48) - Marcos Antônio fala : Eu dou total apoio a Educação Integral, mas, deve-se ressaltar a qualidade, não serve se o jovem ficar lá e o tempo não ser utilizado para adquirir algum conhecimento. (28/11/2012 - 12:50:50) - Giliana Thomé Miola fala : Verdade, Marcos Antonio, a base ruim do fundamental desmotiva o jovem, ele prefere fugir da última etapa da educação básica, foge para um trabalho mau preparado e valorizado.
```

Figura 13: Exemplo de referencia a um usuário em um bate-papo ocorrido no e-democracia

Outra falha identificada foi que nos diálogos não são utilizadas apenas referências a palavras, sinônimos e nomes de usuários, como também outros conectores. A figura 14, que apresenta um exemplo retirado do site e-Democracia, mostra que a participante Giliana utilizou o advérbio "também" para se referir a mensagem anterior.

```
(28/11/2012 - 13:41:51) - Marcos Antônio fala : Giliana para mim a Reforma causará uma transformação. (28/11/2012 - 13:44:36) - Giliana Thomé Miola fala : Também penso assim, uma coisa não exclui outra.
```

Figura 14: Exemplo de bate-papo ocorrido na ferramenta e-democracia

Como mostram os testes que serão apresentados, posteriormente, na seção 3.3.2.4, o algoritmo não teve uma boa acurácia apenas verificando repetições. Portanto, para evoluir o algoritmo descrito é possível utilizar outras características das mensagens em vez de verificar apenas a repetição de palavras. A implementação dessa versão é descrita abaixo.

## 3.3.2.3 Implementação da Segunda Versão

Analisando os diálogos produzidos na ferramenta e-Democracia foi possível perceber que além da repetição de palavras muitos outros padrões são utilizados pelos participantes como referência a uma mensagem anterior. Assim, nessa versão do algoritmo, diversos padrões serão testados para relacionar uma mensagem com a referência anterior. Portanto, o problema se resume a selecionar a mensagem anterior que tem maior probabilidade de ser par de adjacência da mensagem testada a partir de uma série de características das duas mensagens.

De acordo com a descrição do problema, nessa versão do algoritmo foi utilizado um algoritmo de classificação estatística existente para desenvolver um modelo que verifica características que, de acordo com essa pesquisa, são importantes para o encadeamento da discussão. O algoritmo é aplicado em dois momentos, no treinamento, utilizando dados previamente extraídos e classificados nessa pesquisa e na identificação de associações, em que foi desenvolvido um algoritmo que é executado no decorrer do diálogo para converter

as mensagens em atributos que serão utilizados pelo modelo desenvolvido para testar se as mensagens são pares de adjacência.

A seleção das características que serão usadas para a classificação tem um grande impacto no sucesso da mesma. No entanto, geralmente, esse processo é feito por tentativa e erro, de acordo com a intuição do desenvolvedor, que seleciona os atributos que parecem ser relevantes para a solução do problema. Portanto, as características que foram definidas como relevantes nesse trabalho foram aplicadas em um algoritmo estatístico pré-existente para gerar um modelo estatístico que será usado para encadear o diálogo.

Em processamento de linguagem natural, diversas pesquisas foram desenvolvidas para resolver o problema de classificação de sequências de informação. Dentre os algoritmos generativos, a ideia principal dos modelos ocultos de Markov (HMM) (RABINER 1989) é representar uma variável aleatória como uma Cadeia de Markov (ou processo de Markov), com a propriedade particular de que tal variável não pode ser medida diretamente mas sim através de observações possíveis a partir de cada estado da variável. Fundamentados na teoria de Bayes (GHAHRAMANI 1998), são especialmente adequados a situações em que dados observáveis são fartos, sem que haja uma teoria formal conhecida a respeito desses dados. Dentre os algoritmos discriminativos, o modelo de máxima entropia (MANNING; SCHÜTZE 1999) permite a integração de fontes heterogêneas para classificação. Além disso, é interessante que o problema seja solucionado diretamente sem a definição de um modelo mais geral como etapa intermediária.

Segundo (SHRIBERG et al, 2004), um algoritmo que seleciona a mensagem imediatamente anterior à mensagem em que o par de adjacência está sendo buscado tem 79,8% de chances de acerto em bate-papos com dois participantes. Isso acontece, pois, geralmente, os dois usuários interagem como em um diálogo por voz. Porém, em bate-papos com mais de dois usuários interagindo, existe uma grande probabilidade de que outras mensagens sejam inseridas entre a mensagem e o seu par de adjacência.

Assim, nosso problema é selecionar, dentre n possíveis mensagens {s1,...,sn}, a mensagem que maximiza a distribuição de probabilidade condicional (SHRIBERG et al, 2004). Portanto, é possível utilizar um modelo de máxima entropia representar o problema. Para treinar o modelo, Shriberg e colegas [2004] usaram características divididas em 4 categorias: estrutural, de duração, léxica e diálogo.

As categorias citadas acima foram utilizadas em nossa abordagem, com algumas alterações. As características de duração não foram utilizadas pois elas só podem ser obtidas em diálogos por voz. Elas contêm informações como a sobreposição de duas mensagens, isto é, se elas foram faladas no mesmo momento. Na categoria diálogo o artigo cita que informações sobre a ação de uma mensagem não trazem melhoras significativas para o algoritmo e são muito difíceis de obter automaticamente. As características léxicas são informações sobre as palavras de cada mensagem. Elas podem conter dados como o número de palavras, a primeira palavra da mensagem, se a mensagem contém o nome de um usuário ou as palavras em comum entre as mensagens. As características estruturais apresentam informações sobre o diálogo, como quantas mensagens existem entre os possíveis pares de adjacência.

Para a seleção das características a serem utilizadas no modelo, os diálogos registrados no e-democracia foram analisados, além da revisão da literatura apresentada na seção 2.2 e a análise das características utilizadas em (SHRIBERG et al, 2004).

As características escolhidas são discutidas abaixo, mostrando a motivação para a utilização de cada uma delas. A mensagem B representa a mensagem atual, para a qual está se buscando a referência. A mensagem A representa a mensagem anterior à mensagem B sendo testada.

#### Número de mensagens entre A e B

O número de mensagens, isto é, a distância entre A e B pode ajudar a determinar se as duas mensagens são pares de adjacência. A partir dos dados de 1936 mensagens anotadas, para treinamento, nesse trabalho foi possível perceber que a maior distância entre pares de mensagens é de quinze mensagens. A tabela abaixo mostra a frequência da ocorrência de cada distância entre as mensagens, que são pares de adjacência.

Tabela 1: Tabela que representa a distância entre as mensagens de treinamento

Distância entre os pares de adjacência	Frequência
0	461
1	244
2	334
3	223
4	139
5	38
6	21
7	8
8	5
9	1
10	4
11	1
15	1

Portanto, percebemos que as menores distâncias ocorrem com maior frequência e é mais provável que o par de adjacência de uma mensagem seja encontrado nas mensagens mais próximas a ela. Dessa maneira, essa informação pode contribuir com o modelo para determinar as associações.

### Número de palavras em A / Número de palavras em B

Essas características foram utilizadas em (SHRIBERG et al. 2004) e serão testadas nesse cenário para identificar se elas contribuem para a acurácia do algoritmo. Como descrito na seção 2.2, em (REBEDEA et al. 2008), foi proposto um algoritmo que identifica mensagens que tem apenas uma palavra, após a retirada de Stopwords, e verifica se elas pertencem a um padrão de concordância ou discordância. Se o teste for positivo, essa mensagem faz referência para a anterior. Dessa forma, essas características podem contribuir para a identificação desse padrão.

#### Número de palavras em A que também existem em B

Essa característica foi selecionada com base na primeira versão do algoritmo desenvolvido. Geralmente, nos diálogos com grupos de participantes, a repetição é utilizada

para destacar a qual mensagem o usuário está se referindo. Um exemplo é mostrado a seguir:

**YESHUA** fala : por que um simbolo na parede é tão assustador? **estsilva** fala : não é assustador , é inconstitucional

Figura 15: Exemplo de repetição em bate-papo ocorrido na ferramenta e-democracia

A utilização do corretor ortográfico e a radicalização permite que a referência seja encontrada se os radicais forem iguais ou mesmo caso a palavra seja escrita incorretamente.

# Número de n-grams<sup>23</sup> presentes em A e B (n entre 4 e 6)

Como vimos a repetição é utilizada para destacar uma referencia, porém as vezes a palavra é citada no plural, em outro gênero ou outra conjugação. Às vezes a palavra não é reduzida pelo algoritmo de radicalização e necessita da comparação usando n-grams (sequências de letras).

luna lika luz fala : Budista. ACapeletto fala : Budismo é uma Filosofia linda....

Figura 16: Exemplo de bate-papo ocorrido na ferramenta e-democracia

Um exemplo são as palavras budista e budismo vistas acima. Nesse caso o algoritmo de radicalização não reduziu as palavras e elas não seriam incluídas na característica que conta o número de palavras que existem nas duas mensagens. Porém, de acordo com o diálogo, essas mensagens deveriam ser detectadas como uma referencia. Identificando os ngrams como sequências de palavras, com n-gram igual a 4 encontramos as sequências budi, com n-gram igual a 5 as sequências budis. Portanto, essa característica adiciona uma probabilidade maior de que essas mensagens sejam referencias.

#### Primeira e última palavra de A/ Primeira e última palavra de B

A primeira e a última palavra podem indicar, por exemplo, que A é um par de adjacência, como uma pergunta, e provavelmente existe um par para essa mensagem após ela. O mesmo para a mensagem B que pode ser identificada como uma resposta. Dessa maneira, esperasse que essa característica identifique, por exemplo, que mensagens

Um n-gram é uma sequência contínua de n itens de uma sequência. Nesse caso buscamos as sequências de letras contidas simultaneamente no texto das mensagens A e B.

começando em palavras como aonde, porque, quando, como e terminando em interrogação representam perguntas.

A contém o nome do usuário que enviou a mensagem B? / B contém o nome do usuário que enviou a mensagem A?

O nome dos usuários presentes no bate-papo é um recurso muito utilizado para indicar que a mensagem é uma resposta ou faz referencia a uma mensagem enviada por aquele usuário. O exemplo abaixo foi retirado de um diálogo sobre o tema médicos estrangeiros no Brasil ocorrido no site e-democracia.

```
(12/06/2013 - 11:03:08) - Rodrigo Franco fala : Por que o ministerio não cria a carreira de estado para médicos e faz concurso pra eles?
(12/06/2013 - 11:03:54) - Vania41 fala : As pessoasprecisam ser atendidas HOJE, não no dia do MUNDO MÀGICO que o CFM quer, com tudo ajeitado, com todos os exames, estrutura, salários fantasiosos... as pessoas precisam ser atendidas: Médicos Estrangeiros e + Provab Já!!! CFM desça para o mundo REAL!!
(12/06/2013 - 11:05:06) - Cezar fala : Vania41, minha querida!! Adianta chamar encanador pra resolver o problema de água no NORDESTE BRASILEIRO???
(12/06/2013 - 11:05:18) - Vania41 fala : Carreira de estado não garante nada Rodrigo Franco...Veja quantos municípios estão sem Juízes, que tem carreira de estado...eles só querem trabalhar na cidade!!! Medicos estrangeiros e + PROVAB já
(12/06/2013 - 11:06:16) - Rodrigo Franco fala : FAÇA CONCURSOS
(12/06/2013 - 11:06:44) - Vania41 fala : Então Cezar, a sua solução é que morram??? Vc já houviu falar em atenção primária a saúde? Promoção à saúde? várias estratégias podem evitar doenças com tecnologias leves.
```

Figura 17: Exemplo de referencia a nomes em um bate-papo ocorrido na ferramenta e-Democracia Abaixo são destacadas todas as características utilizadas para gerar o modelo.

#### Características estruturais:

Número de mensagens entre A e B

#### Características léxicas:

- Número de palavras em A
- Número de palavras em B
- Número de palavras em A que também existem em B
- Número de n-grams presentes em A e B (n entre 4 e 6)
- Primeira e última palavra de A
- Primeira e última palavra de B
- A contém o nome do usuário que enviou a mensagem B?
- B contém o nome do usuário que enviou a mensagem A?

Portanto, o algoritmo implementado utiliza o modelo obtido no treinamento realizado nesse trabalho para classificar que duas mensagens são pares de adjacência ou não. Assim, além de comparar o número de palavras em comum entre as duas mensagens também são verificadas outras características importantes que foram utilizadas para a construção do modelo de máxima entropia.

Assim, a partir da extração de logs de bate-papos e da sua classificação, identificando pares de adjacência, foi desenvolvido um algoritmo que, para cada mensagem, recupera as vinte mensagens anteriores e extrai as características definidas acima entre essas duas mensagens. Para o treinamento do modelo, o algoritmo de máxima entropia utilizado recebe como parâmetro um arquivo de texto com as características extraídas para cada par de mensagens, pelo algoritmo desenvolvido nessa pesquisa, além da indicação, extraída do arquivo de mensagens anotadas, se as duas mensagens são pares de adjacência. Finalmente, o algoritmo de máxima entropia cria um modelo, a partir desses dados, que será usado posteriormente na construção do Comunicografo durante os bate-papos.

#### 3.3.2.4 Testes

Para o treinamento e teste dos algoritmos implementados foram utilizados trinta e dois logs de bate-papos do Gtalk e dezessete audiências públicas coletadas no site e-democracia. Esses diálogos foram transformados em XML de acordo com o esquema XSD construído com base no modelo apresentado na figura 11. Posteriormente, os XML's foram anotados manualmente, por apenas uma pessoa, identificando os pares de adjacência. O número total de mensagens anotadas nos logs foi de 1936.

O primeiro algoritmo descrito foi utilizado para analisar os logs, identificar mensagens em que aparece a mesma palavra e classificar como uma referência. Portanto, após a classificação do algoritmo, as anotações manuais sobre os pares de adjacência foram recuperadas para descobrir a acurácia do algoritmo. Assim, a acurácia foi calculada como o número de mensagens que foram anotadas como pares de adjacência e foram identificadas como tal pelo algoritmo sobre o número total de pares de adjacência anotados. Com base nesse cálculo, a acurácia do algoritmo foi de 635/1480, correspondendo a aproximadamente 42% de acerto.

64

Um exemplo de recuperação de referência, após o tratamento das mensagens,

utilizando esse algoritmo é mostrado abaixo:

Usuario1: tem que **marc** outr encontr que nem aquel iguatem

Usuario2: vam marc sim vm cinem tb

Devido a esse baixo desempenho, foi desenvolvido um novo algoritmo utilizando um modelo de máxima entropia. Esse modelo permite que sejam usadas diversas características além da repetição de palavras entre as mensagens. Para treinar o modelo foi utilizada a API

do Apache OpenNLP<sup>24</sup> para o algoritmo de máxima entropia. A API descrita foi escolhida

pela facilidade de uso e por utilizar técnicas para reduzir o tamanho dos modelos tornando

mais rápida a sua leitura.

O objetivo do modelo de máxima entropia é integrar informações heterogêneas de

diversas fontes para classificação. O algoritmo utiliza diversas características que devem ser

descritas pelo experimentador. Elas são escolhidas de acordo com o conhecimento deste

sobre os dados, representando as características mais importantes para a sua classificação.

Portanto, para o treinamento do modelo, os logs de bate-papos extraídos foram

utilizados, totalizando 1290 mensagens, representando aproximadamente 2/3 dos dados

coletados. O algoritmo desenvolvido percorre os logs, recupera os pares de adjacência

anotados e as características, descritas no capitulo anterior, para cada par buscado a partir

das vinte mensagens anteriores a cada mensagem. Essas características são gravadas em um

arquivo de texto que será lido para criar o modelo de máxima entropia.

Finalmente, os logs de bate-papos restantes foram utilizados para testar o modelo

gerado. Dessa maneira, a execução do algoritmo foi simulada, percorrendo, para cada

mensagem do log, as vinte mensagens anteriores e usando o modelo criado no treinamento

para descobrir a probabilidade delas serem pares de adjacência. De acordo com os testes

realizados, o novo algoritmo classificou 1381 mensagens que eram pares de adjacência

corretamente, obtendo uma acurácia de 1243/1480, correspondendo a 84,98%,

significativamente melhor que o algoritmo anterior.

<sup>24</sup> OpenNLP. Disponível em <a href="http://opennlp.apache.org/">http://opennlp.apache.org/</a> 10/04/2013

## 3.3.2.5 Identificação de Dúvidas

Os diálogos produzidos em ferramentas de comunicação textual, com vários participantes, são confusos e não seguem uma linearidade. Tal fato decorre da quantidade de usuários e a velocidade em que as mensagens são digitadas fazendo com que vários assuntos sejam discutidos ao mesmo tempo tornando complicada a tarefa de acompanhar o diálogo. Portanto, fica difícil determinar quem está falando com quem sobre o quê (PIMENTEL et al. 2003).

Para que seja possível identificar mensagens que representam dúvidas, os logs anotados foram analisados e foram retiradas as palavras mais utilizadas para representar que o não entendeu algum conceito.

A implementação do algoritmo utiliza uma lista com as palavras mais comuns para identificar as mensagens que representam dúvidas. Essas palavras foram extraídas a partir dos diálogos coletados para treinamento.

Assim, essas mensagens são marcadas no Comunicografo para que o módulo de recomendação, posteriormente, considere essa informação para recuperar conteúdo interessante que contribua com o diálogo.

#### 3.3.3 Módulo de Recomendação de Conteúdo

O módulo de recomendação tem o objetivo de agregar informações construtivas a discussão utilizando dados obtidos a partir da análise do diálogo realizada no módulo anterior. Para alcançar esse fim é necessário que, a partir do Comunicografo construído, das dúvidas encontradas e da mensagem enviada, sejam obtidas informações relevantes que complementem as mensagens enviadas.

Portanto, esse módulo é divido em três passos (figura 18): identificação de palavras relevantes na mensagem, construção de consultas que serão executadas nos datasets da Web de Dados e desambiguação dos resultados obtidos a partir das consultas construídas.

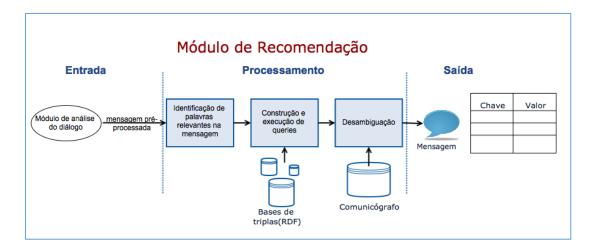


Figura 18: Diagrama do módulo de recomendação de conteúdo

Para extrair informações relevantes da Web de Dados é necessário que sejam formuladas consultas que possam identificar dados relacionados com o diálogo. Para isso, inicialmente, o módulo de recomendação analisa a mensagem enviada e identifica palavras que, possivelmente, representam conceitos específicos, principalmente nomes próprios, que se referem a entidades descritas nos datasets utilizados. Esse passo de identificação de palavras e conjunto de palavras relevantes na mensagem foi dividido em três algoritmos. O primeiro algoritmo utiliza a extração de entidades nomeadas para identificar palavras chaves no bate-papo. O segundo é baseado em um algoritmo de análise morfológica para identificar palavras ou grupos de palavras, que não foram encontradas pelo algoritmo anterior e que podem ser relevantes para a construção das consultas. O terceiro, executado no módulo anterior, recupera palavras que representem dúvidas para que sejam recomendadas informações que solucionem a dúvida do participante.

O reconhecimento de entidades nomeadas é uma técnica da área de processamento de linguagem natural que tem como objetivo a identificação de nomes próprios e conceitos como pessoas, lugares e organizações. A classificação de entidades é muito útil na análise semântica das frases, identificando o significado de palavras ou grupos de palavras ambíguos. Portanto, o AidChat utiliza um algoritmo, já existente com pequenas alterações, para identificar as entidades nomeadas presentes na mensagem enviada e busca, nos datasets, informações que descrevam as entidades.

Porém, o algoritmo de reconhecimento de entidades nomeadas pode não reconhecer alguns conceitos. Portanto, após o reconhecimento de entidades é feita a análise

morfológica da mensagem com o objetivo de encontrar palavras que também são candidatas à recomendação e não foram encontradas pelo algoritmo de entidades nomeadas. Essa análise utiliza um algoritmo já existente que, baseando-se no contexto e significado, categoriza as palavras em classes gramaticais. Assim, o algoritmo identifica nomes, endereços, nomes próprios e realiza consultas nas bases de dados, retornando informações que descrevam essas palavras.

Finalmente, caso a mensagem contenha palavras que representem dúvidas, identificadas no módulo anterior, as mensagens às quais elas se referem são recuperadas. Isto é, o algoritmo recupera a mensagem à qual ela faz referencia, a referencia dessa mensagem e assim por diante. Assim o algoritmo desenvolvido utiliza as mensagens referenciadas, isto é, o conjunto de mensagens que se referem àaquele assunto, para formular uma consulta em que sejam retornadas informações que ajudem a solucionar a dúvida do participante.

A partir das consultas formuladas e executadas na Web de Dados, é gerada uma estrutura chave/valor que armazena a(s) palavra(s) utilizada(s) na busca e a informação encontrada. Assim, uma mesma palavra não é consultada mais de uma vez e os dados são armazenados para serem apresentados para os usuários após a análise e desambiguação.

Antes de apresentar as informações aos usuários é importante verificar se elas são realmente relevantes de acordo com o diálogo, evitando assim equívocos nos algoritmos anteriores e no contexto das informações recuperadas. Além disso, as consultas executadas anteriormente também podem retornar mais de um resultado para cada entidade extraída. Portanto, é necessário executar a desambiguação dos resultados retornados, isto é, identificar a qual entidade existente nas bases consultadas a palavra se refere de acordo com o contexto sendo discutido. Para isso é feita uma comparação da similaridade semântica, a partir de um algoritmo existente, do texto candidato a recomendação com a mensagem enviada. Além disso, também é calculada a similaridade entre esse texto e as entidades nomeadas previamente recuperadas no grafo que contém a hierarquia das mensagens referenciadas pela ultima mensagem enviada. Dessa forma é possível validar se a recomendação faz sentido no contexto ao qual a mensagem faz parte.

Finalmente, a partir da definição de informações interessantes para o diálogo, em conjunto com a mensagem enviada é necessário definir como essas recomendações serão apresentadas para os participantes. Para que o diálogo não seja interrompido e o usuário possa decidir quais informações ele deseja visualizar, as palavras são apresentadas no canto esquerdo da tela na forma de ligações. Essas palavras representam a chave do mapa chave/valor construído anteriormente. Assim, caso o usuário tenha dúvidas, ou possua pouco conhecimento em relação ao contexto da palavra ele pode, rapidamente, passar o ponteiro do mouse acima do link (palavra) e visualizar a informação ali contida (figura 19).



Figura 19: Exemplo de recomendação no bate-papo implementado

#### 3.3.3.1 Implementação

O módulo de recomendação recebe como entrada a última mensagem enviada no bate-papo. Essa mensagem será analisada em conjunto com o Comunicografo, gravado anteriormente, produzindo como saída informações complementares que serão apresentadas para os usuários em conjunto com a mensagem enviada.

As informações obtidas por esse módulo são recuperadas através de bases de dados pré-definidas que contém informações no contexto do cenário abordado no bate-papo.

A motivação para o uso de bases da Web de Dados é a abrangência de contextos, o constante aumento de bases de dados publicadas dessa forma e a interoperabilidade entre elas. Com o aumento das bases publicadas e as ligações entre elas a abordagem pode contemplar uma gama maior de contextos que possibilita a sua aplicação para um conjunto maior de cenários. Além disso, esse formato de publicação de dados permite que as

informações geradas pela ferramenta estejam preparadas para a publicação na núvem de WoD.

Nessa pesquisa, devido à definição de um cenário específico foram escolhidos datasets específicos para esse cenário. Os datasets escolhidos foram a Dbpedia que armazena dados da Wikipédia, pela abrangência de contextos e o LexML, uma base de dados que contém dados da legislação brasileira.

Como discutido no capítulo anterior, inicialmente, a mensagem é analisada com o objetivo de recuperar palavras que representem conceitos específicos que serão usados na construção de consultas para extrair informações relevantes dos datasets.

Para recuperar palavras relevantes na mensagem, três algoritmos foram implementados. No primeiro foi utilizado um algoritmo já existente que recupera entidades nomeadas de acordo com a frase. O algoritmo usado foi o DBpedia Spotlight (DAIBER et al. 2013) pois ele retorna o recurso da Dbpedia ao qual a entidade encontrada se refere. O resultado da extração de entidades é armazenado para que posteriormente as consultas sejam executadas. O DBpedia Spotlight retorna o recurso da DBpedia PT, pois a versão em português do algoritmo foi utilizada, correspondente a entidade encontrada. Portanto, devido à falta de informações de resumos na versão da DBpedia em português, foi necessário alterar o algoritmo para retornar o recurso correspondente na DBpedia em inglês. A alteração construiu um SPARQL que recupera o recurso da DBpedia EN a partir do recurso da DBpedia PT utilizando a propriedade owl:sameAs. Essa propriedade é utilizada para identificar que duas URIs fazem referencia ao mesmo conceito.

Após a extração de entidades foi verificado que alguns conceitos relevantes não foram encontrados, principalmente palavras que representam conteúdos de legislação. Portanto, a API do CoGrOO (SILVA 2013) para análise morfológica foi usada para recuperar nomes próprios e numerais que complementam as entidades recuperadas anteriormente.

Para aprimorar a recuperação de dados da legislação foi desenvolvido um algoritmo que utiliza uma lista de palavras e grupos de palavras que representam dados de legislação e recupera entidades e numerais posteriores a essas palavras, utilizando a API do CoGrOO. A lista de palavras descrita abaixo foi definida utilizando o LexML e a análise de palavras utilizadas nos bate-papos extraídos do e-Democracia.

- Lei
- Lei complementar
- Lei constitucional
- Lei delegada
- Lei Orgânica
- Portaria
- Enunciado
- Deliberação
- Regimento Interno
- Resolução
- · Projeto de lei
- PL
- Constituição
- Emenda
- Norma
- Decreto
- Projeto de emenda constitucional
- PEC

(19:42:40) nolasco: a pec 71 diminui os custos?

Figura 20: Exemplo retirado de bate-papo ocorrido na experimentação

No exemplo apresentado na figura 20 o usuário faz uma pergunta sobre uma PEC (projeto de emenda constitucional) existente na legislação brasileira. Nesse caso, o algoritmo desenvolvido encontraria a palavra PEC e , a partir da API do CoGrOO, o numeral 71 que se encontra após a palavra PEC. Assim, a abordagem armazena o conjunto de palavras PEC 71 para usar posteriormente na construção das consultas.

O terceiro algoritmo desenvolvido para recuperar palavras relevantes apenas verifica no Comunicografo se a mensagem enviada contém alguma palavra que seja referente a uma dúvida. Caso exista, o algoritmo recupera a lista de mensagens às quais ela faz referencia, isto é, a thread de mensagens do assunto e utiliza as entidades recuperadas anteriormente nessas mensagens para formular a consulta posteriormente.

Após a execução de cada um dos três algoritmos anteriores, um algoritmo desenvolvido identifica se o conceito chave encontrado já havia sido recuperado anteriormente. Esses conceitos são removidos da lista para que eles não sejam buscados mais de uma vez nas consultas executadas posteriormente.

A partir da extração das palavras relevantes para a construção das consultas que serão executadas nos *datasets* escolhidos foram desenvolvidos dois algoritmos. Foi necessário desenvolver dois algoritmos diferentes, pois a base de dados do LexML está no formato XML e a Dbpedia está no formato RDF.

O algoritmo desenvolvido para a DBpedia tem como entrada a lista de palavras ou grupos de palavras encontrados pelos algoritmos anteriores. Como saída são retornados um mapa chave/valor que contém a palavras(s) chave usadas na busca e a recomendação encontrada para ela(s). Para que a recomendação seja buscada foi necessário definir o atributo da DBpedia que seria apresentado aos participantes, pois cada recurso contém diversos atributos que o descrevemna base de dados (figura 21).

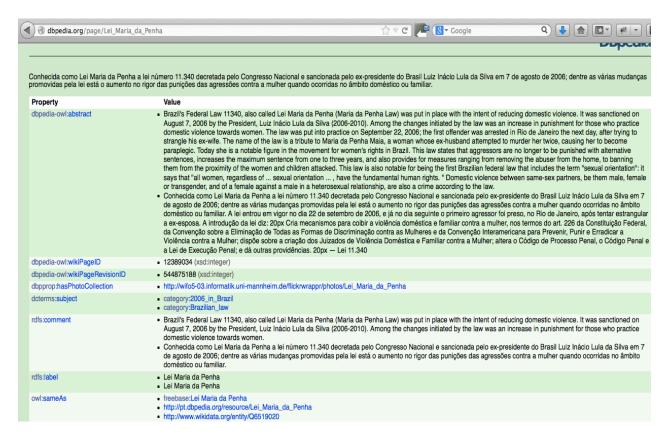


Figura 21: Informação sobre a Lei Maria da Penha disponível na Dbpedia

Na DBpedia podem ser encontrados atributos que contém pequenas descrições como o dbpedia-owl:abstract, que é um resumo mais longo, e o rdfs:comment, que é um resumo mais curto. Portanto, o resumo contém uma descrição básica e sucinta da entidade e foi escolhido como recomendação a ser apresentada para os usuários.

Assim, o algoritmo percorre a lista de palavra(s) chave(s) e executa as consultas desenvolvidas. Caso a chave seja um recurso da DBpedia, encontrado pelo DBpedia Spotlight, é executado um SPARQL que busca exatamente esse recurso. Caso a chave seja um literal, o SPARQL desenvolvido busca todos os recursos que contém o literal como *label* e armazena no mapa chave-valor que contém as recomendações.

O LexML disponibiliza um serviço Web que segue o padrão SRU<sup>25</sup> em que sua base de dados pode ser consultada. Assim, o algoritmo desenvolvido processa a lista de chaves armazenadas para busca e executa consultas buscando dados da legislação em que o título contenha a chave. Assim como a Dbpedia, o LexML também disponibiliza um resumo curto e sucinto da informação.

A partir do mapa de recomendações encontradas é necessário verificar se elas são realmente relevantes para o contexto sendo discutido. Para isso, foi implementado um algoritmo que utiliza uma API pré-existente, que calcula a similaridade semântica entre textos. A partir de cada consulta executada, o algoritmo recupera o mapa armazenado com as palavras chaves e suas respectivas recomendações. Para as consultas que retornaram mais de um resultado, as recomendações são ordenadas de acordo com a similaridade com o diálogo e nas consultas que retornaram apenas um resultado também é calculada a similaridade da recomendação com o diálogo.

Para identificar a similaridade do texto a ser recomendado com o diálogo, o algoritmo compara a mensagem enviada com o texto recomendado e recupera as mensagens referenciadas pela mensagem enviada e utiliza as entidades nomeadas e nomes próprios contidos nelas para a comparação com o texto a ser recomendado. O algoritmo faz a média a partir das duas similaridades encontradas e ordena as recomendações.

Na figura 22, que destaca um trecho do Comunicografo gerado na experimentação,

\_

<sup>&</sup>lt;sup>25</sup> SRU Search/Retrieval via URL. Disponível em: <a href="http://www.loc.gov/standards/sru/">http://www.loc.gov/standards/sru/</a>>.17/05/2013

caso fosse encontrada uma recomendação para a última mensagem, da participante *usuario2*, o algoritmo desenvolvido, compararia o texto da recomendação com essa mensagem, além de comparar o texto recomendado com as entidades nomeadas contidas na thread marcada em vermelho. Finalmente, é feita a média das duas similaridades para obter a similaridade da recomendação com o diálogo.

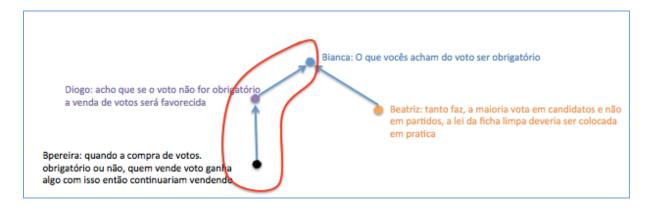


Figura 22: Trecho do Comunicógrado gerado na experimentação

A API utilizada para a comparação da similaridade semântica foi a ESA Semantic Relatedness<sup>26</sup>. A API calcula a similaridade entre dois fragmentos de texto baseada no método Explicit Semantic Analysis (ESA) (GABRILOVICH; MARKOVITCH et al. 2007) que utiliza propriedades semânticas descritas no corpus da Wikipedia e se apoia na hipótese de que palavras que ocorrem em contextos similares tendem a possuir significados similares. Para eliminar os termos menos correlacionados foi definido um valor de corte (threshold) de 0.0005.

Finalmente, o algoritmo desenvolvido retorna o mapa de recomendações para a tela. Assim, os usuários podem visualizar as palavras chaves e obter informações sobre elas durante o diálogo.

## 3.3.3.2 Bases de Dados Utilizadas

Esta seção descreve as bases de dados que foram utilizadas pelo módulo de recomendação, descrito na seção anterior, como fontes para recuperar as informações que serão mostradas aos usuários do bate-papo.

-

<sup>&</sup>lt;sup>26</sup> ESA Semantic Relatedness. Disponível em: <a href="http://www.amtera.com.br/webapi.html">http://www.amtera.com.br/webapi.html</a>>.12/05/2013

#### **3.3.3.2.1** LexML

O LexML utiliza um modelo de referencia que define as entidades utilizadas para representar os dados de sua base. Esse modelo é baseado na ontologia FRBROO (Functional Requirements for Bibliographical Records) (DOERR et al. 2008). "Uma das principais vantagens da adoção desse modelo é a percepção dos diversos níveis de abstração dos recursos de informação. " Assim podemos representar os metadados de cada recurso, como por exemplo, a criação e modificações de uma norma jurídica.

O projeto utiliza nomes uniformes de recursos - Uniform Resource Names (URNs) para fornecer identificadores unívocos para cada objeto do banco de dados. Assim, os documentos jurídicos podem se referir a outros e estes podem ser acessados imediatamente.

Foi definido um esquema XML para a estruturação dos textos de normas, julgados e projetos de normas. Devido à falta de padronização da técnica legislativa em relação às três esferas e à existência de normas federais com articulação incompatível com a Lei Complementar nº 95/1998 (principalmente normas anteriores a 1998), foram criados dois esquemas, um flexível e um rígido.

Ao se definir uma nova linguagem na notação XML é comum surgirem dúvidas quanto à escolha entre o construtor elemento e atributo em uma determinada situação. Utilizou-se a seguinte diretriz no LexML: o conteúdo textual dos elementos deve ser suficiente para reproduzir o texto constante do documento oficial (em papel ou arquivo digitalmente assinado) na ordem em que ele se apresenta; os atributos codificam informações de controle do conteúdo texto, tais como identificadores, referências, indicação da língua etc.

Todos os documentos LexML compartilham o mesmo elemento raiz <LexML>, cujo conteúdo inicial é obrigatoriamente o elemento <Metadado> seguido do elemento que identifica o tipo de documento. A utilização de um único elemento raiz segue o Design Pattern "Universal Root".

O LexML prevê inicialmente os seguintes tipos de documentos:

Norma – Constituição, Emendas, Leis, Decretos etc.

ProjetoNorma – Proposições legislativas.

Jurisprudencia – Súmulas e Acórdãos.

DocumentoGenerico – Para outros tipos de documentos.

Anexo – Anexos dos documentos acima.

Os elementos que definem a hierarquia de um documento articulado, de acordo com a Lei Complementar nº 95/1998, são apresentados no arquivo "lexml-br-rigido.xsd". Inicialmente são apresentados os elementos que agrupam artigos (Parte, Livro, Titulo, Capitulo, Secao e Subsecao), em seguida a unidade básica de articulação (Artigo), e, por fim, os dispositivos de artigo (Caput, Inciso, Alinea, Item, Paragrafo). Os elementos correspondentes, no esquema flexível, são apresentados no arquivo "lexml-flexivel.xsd". Os esquemas rígido e flexível fazem inclusão do esquema base, localizado no arquivo "lexml-base.xsd".

Na seqüência, são apresentados os elementos genéricos correspondentes a cada um dos cinco patterns de modelo de conteúdo: Agrupamento (Container), Agrupamento Hierarquico (Hierarchy), Bloco (Block), Em Linha (Inline), Marca (Marker). Ao utilizar estes elementos genéricos é necessário especificar o atributo "nome" significativo.

Em seguida, o esquema apresenta os elementos "importados" do vocabulário HTML que tratam da codificação do texto corrido, de listas, de tabelas e de outros elementos de formatação tais como negrito, itálico, sublinhado, subscrito e sobrescrito. Trata-se de uma simplificação do vocabulário HTML.

Por fim, são apresentados os elementos de metadados que tratam de vários aspectos, tais como identificação, contexto, indexação, ciclo de vida e informações sobre a publicação.

O controle de vocabulários desempenha papel fundamental na organização da informação legislativa e jurídica. Ele permite, entre outras coisas, a realização de pesquisas sofisticadas, a execução do processo de normalização da URN e a integração dos vocabulários utilizados pelas diversas casas.

## 3.3.3.2.2 DBpedia

A Wikipedia<sup>27</sup> é o sexto *site* mais usado, a enciclopédia mais consultada e tornou-se uma das fontes de conhecimento centrais no ciberespaço, mantida por milhares de contribuintes. Ela já possui versões oficiais em 287 línguas e é mantida colaborativamente por usuários da Web.

O projeto DBpedia aproveita essa enorme fonte de conhecimento extraindo dados estruturados da Wilkipedia em 119 idiomas e disponibilizando-os na Web. A DBpedia extrai essa informação na forma de triplas RDF fazendo com que as entidades existentes na Wikipedia sejam descritas de forma semântica e consultas mais expressivas possam ser construídas sobre esses dados.

O núcleo do DBpedia consiste em um processo de extração de infoboxes. Infoboxes são modelos da Wikipedia que geralmente são exibidos no canto superior direito de artigos. Além desse tipo de extração, a DBpedia também conta com 19 extratores que processam os conteúdos da Wikipedia como rótulos, imagens, georreferencias, ligações etc. Os extratores são executados para manter a DBpedia atualizada de acordo com a Wikipedia.

Dessa forma, os dados da DBpedia são disponibilizados tanto para download quanto para consulta online através de um Sparql Endpoint. Um Sparql Endpoint é um serviço Web que permite que consultas Sparql sejam realizadas em bases de triplas.

O conjunto de dados da DBpedia apresenta rótulos e resumos para 12,6 milhões de coisas únicas em 119 línguas diferentes ; 24,6 milhões de ligações para imagens e 27,6 milhões de ligações para páginas Web externas; 45,0 milhões de ligações externas para outros conjuntos de dados RDF , 67,0 milhões de ligações para categorias Wikipedia , e 41,2 milhões de categorias YAGO.

A partir da publicação dos dados da DBpedia diversas outras bases de dados em RDF, focados em diversos contextos, foram sendo publicadas e ligadas a DBpedia. Portanto, além de possuir conhecimento sobre diversos domínios, também é possível obter informações mais específicas de outras bases de dados através de suas ligações.

-

Wikipedia. Disponível em: <a href="http://www.wikipedia.org/">http://www.wikipedia.org/</a>>. 02/01/2014

Além disso, a DBpedia também contém uma das maiores ontologias multi-domínios que existe. A ontologia é publicada em conjunto com os dados e também é atualizada automaticamente, de acordo com as atualizações da Wikipedia.

Portanto, a DBpedia armazena uma grande quantidade de entidades que abrangem diversas áreas do conhecimento. Isso fez com que ela se tornasse um *hub* central da WoD devido a quantidade de datasets que foram ligados aos seus conceitos.

#### 3.4 Problemas Encontrados

Ao longo da implementação e da experimentação da abordagem AidChat diversos problemas foram encontrados.

Uma das grandes dificuldades encontradas foi o desempenho da ferramenta desenvolvida para recomendações. A quantidade de consultas *online* executadas ao SPARQL Endpoint da DBpedia, ao serviço Web do LexML e ao serviço Web do algoritmo ESA Semantic Relatedness são grandes contribuintes para que o desempenho não seja bom. Portanto, foi percebido que é necessário reduzir a quantidade de consultas feitas *online*. Para isso é possível fazer o *download* da base de dados da DBpedia e obter o algoritmo ESA Semantic Relatedness para consultas locais.

Outra questão que torna o desempenho do sistema apoiado na abordagem lento é a construção do Comunicografo. O algoritmo que o constrói percorre, para cada mensagem, as vinte mensagens anteriores e extrai um conjunto de características definidas no modelo para cada par de mensagens. Assim, o algoritmo testa cada par de mensagens, entre a mensagem enviada e as anteriores, de acordo com o Comunicografo, e as classifica como par de adjacência ou não. Quando o algoritmo encontra um par de adjacência para a última mensagem enviada ele interrompe o processo. Portanto, a quantidade de processamento necessário é muito grande. Para diminuir esse problema é possível permitir que os participantes ajudem no encadeamento do diálogo selecionando uma referencia explicita ao enviar uma mensagem.

Outro problema foi o formato da base de dados do LexML, disponível em XML. Também não foi possível obter o *download* da base completa para transformá-la em RDF. Assim, foi necessário utilizar o serviço Web disponibilizado para executar consultas

empregando a linguagem Contextual Query Language (CQL)<sup>28</sup>. Assim, as consultas são menos ricas semanticamente. Além disso, o modelo definido para representar a base é muito simples e também impossibilita consultas mais semânticas, faltando, por exemplo, que as leis sejam organizadas em categorias que representem contextos.

Outra dificuldade, com relação às bases de dados utilizadas, foi encontrada na DBpedia em português. A API Dbpedia Spotlight é utilizada nessa pesquisa na sua versão para língua portuguesa e retorna URIs de recursos existentes na DBpedia PT. Porém, a versão em português da DBpedia não contém resumos, fazendo com que seja necessário desenvolver uma consulta SPARQL que seja executada na versão da DBpedia em inglês para retornar o recurso correspondente a partir da propriedade owl:sameAs. Assim, utilizando o recurso correspondente é possível recuperar o resumo de sua descrição.

Outro problema, identificado a partir do diálogo realizado na ferramenta durante a experimentação, foi que algumas entidades recomendadas eram pouco relevantes para o contexto sendo discutido. Essas entidades representam conceitos muito comuns, amplamente conhecidos, que não proporcionam aprendizado e atrapalham a visualização de outras entidades mais construtivas para o diálogo.

Devido à velocidade com que as interações ocorrem no bate-papo, principalmente quando o número de participantes é muito grande, é necessário que os participantes tenham muita atenção ao diálogo e interajam rapidamente. Dessa maneira, para que as recomendações sejam construtivas para o diálogo elas devem ser de fácil visualização. Portanto, recomendações irrelevantes podem atrapalhar a visualização do participante fazendo com que ele perca tempo tentando encontrar a entidade que corresponde a sua dúvida.

Portanto, é necessário discutir como eliminar entidades que não são interessantes no contexto do diálogo. Nessa pesquisa, uma tentativa de diminuir recomendações fora do contexto foi feita calculando a similaridade semântica entre o diálogo e o texto a ser recomendado.

A ferramenta DBpedia Spotlight permite que sejam selecionados tipos de entidades que serão extraídas do texto (figura 23). Para isso, os atributos existentes nas ontologias das

.

<sup>&</sup>lt;sup>28</sup> CQL – The contextual query language. Disponível em: <a href="http://www.loc.gov/standards/sru/cql/index.html">http://www.loc.gov/standards/sru/cql/index.html</a>. 04/07/2013

bases de dados utilizadas pela ferramenta são mostrados para o usuário e ele seleciona os tipos de entidades que são relevantes para o contexto do texto que serão extraídas entidades nomeadas.

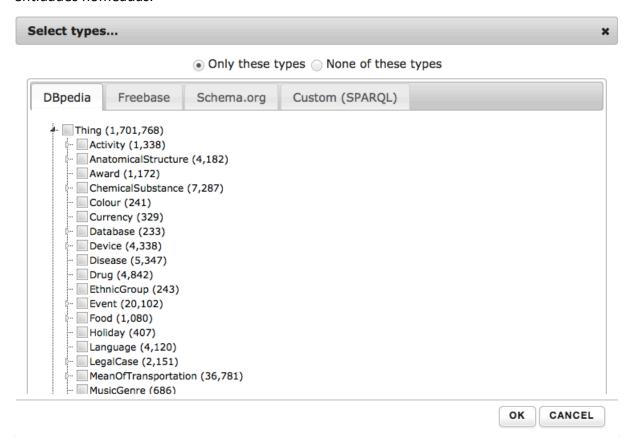


Figura 23: Seleção de tipos de entidades na DBpedia Spotlight

Portanto, uma opção para filtrar os tipos de entidades que serão recomendadas seria permitir que o usuário configurasse os tipos de entidades relevantes para o contexto que será discutido no bate-papo.

A análise desses problemas, identificados a partir da experimentação ou durante o desenvolvimento, permite que sejam discutidas estratégias que podem diminuir o impacto causado por eles, ajudando a definir trabalhos futuros a essa pesquisa.

## 4 Avaliação

Para a avaliação da abordagem proposta foram realizados dois experimentos em sessões de bate-papo, utilizando a ferramenta desenvolvida, em que participantes foram estimulados a dialogar sobre determinado assunto. Ao longo do bate-papo os participantes puderam acompanhar as recomendações realizadas pela abordagem. Portanto, foi realizada uma pesquisa qualitativa com caráter exploratório estimulando o entrevistado a pensar e se expressar livremente para avaliar a abordagem de recomendação.

Após a participação no bate-papo, os participantes foram incentivados a responder um questionário avaliando a abordagem e as recomendações propostas durante o diálogo. Questionários são uma forma rápida e simples para avaliar as opiniões, objetivos, anseios e preferências de pessoas. O questionário reuniu questões de resposta aberta e fechada. As questões de resposta fechada foram estruturadas de maneira que os participantes marcavam as respostas de acordo com a escala de Likert de sete níveis (LIKERT 1932). Essa escala é muito utilizada na construção de questionários e tem o objetivo de registrar a discordância ou concordância do participante com dada afirmação.

Finalmente, o diálogo realizado na experimentação foi analisado, bem como o Comunicografo gerado pela ferramenta durante o diálogo e os questionários respondidos pelos participantes.

## 4.1 Implementação no Cenário Escolhido

O cenário escolhido para a validação da proposta desenvolvida foi o portal do e-Democracia. O portal, desenvolvido pela Câmara dos Deputados, tem a proposta de incentivar a participação da sociedade no debate de temas importantes para o país. Segundo os autores do projeto, o envolvimento dos cidadãos na discussão de novas propostas de lei contribui para a formulação de políticas públicas mais realistas e implantáveis.

A escolha desse cenário se baseou na estrutura dos bate-papos que ocorrem no portal. Devido à seriedade dos assuntos discutidos a escrita é mais formal, com menos erros ortográficos e abreviações. Outra motivação é o tema das discussões, a legislação brasileira, que é muito extenso e complexo, pela grande quantidade de leis e projetos de leis e termos

jurídicos utilizados, causando dificuldades para o entendimento e debate em conjunto com a população.

O portal e-Democracia é dividido em comunidades temáticas que discutem diversos assuntos. Em cada uma delas é possível participar de debates em fóruns e bate-papos em tempo real entre os membros da comunidade e deputados envolvidos com o tema. Elas também disponibilizam wikis em que os membros podem propor alterações aos projetos de lei em discussão e bibliotecas virtuais onde são armazenados documentos digitais relacionados com o tema como estudos, artigos e leis existentes que tratam do tema. Além disso, os usuários cadastrados também podem utilizar um bate-papo para a comunicação entre eles.

A grande contribuição do portal é permitir que os parlamentares envolvidos com a matéria acompanhem as discussões realizadas pela população e as considerem para auxiliar nas suas decisões. Assim, a população pode participar ativamente das assembleias legislativas podendo até enviar perguntas aos deputados.

Infelizmente, por problemas burocráticos para a liberação do código do *portlet* de bate-papo desenvolvido pela equipe do e-Democracia, não foi possível implementar a abordagem proposta em seu ambiente original. Entretanto, o objetivo é que a integração da abordagem proposta com o portal do e-Democracia seja feita no futuro.

Dessa maneira, para testar a abordagem em um ambiente similar ao e-Democracia, ela foi implementada utilizando um servidor de mensagens instantâneas e um cliente XMPP que permite bate-papos entre grupos de usuários. O servidor utilizado foi o Openfire<sup>29</sup>, um servidor simples de instalar e administrar, que permite que sejam criadas salas de bate-papo e usuários, além de permitir a gravação das mensagens enviadas. Como cliente foi utilizado o SparkWeb<sup>30</sup>, uma versão para Web do cliente de mensagens instantâneas Spark<sup>31</sup>. O

<sup>&</sup>lt;sup>29</sup> Ignite Realtime – Openfire. Disponível em: <a href="http://www.igniterealtime.org/projects/openfire/">http://www.igniterealtime.org/projects/openfire/</a>.
17/10/2013

<sup>30</sup> Ignite Realtime – SparkWeb. Disponível em: <a href="http://www.igniterealtime.org/projects/sparkweb/index.jsp">http://www.igniterealtime.org/projects/sparkweb/index.jsp</a>>. 17/10/2013

<sup>31</sup> Ignite Realtime – Spark. Disponível em: <a href="http://www.igniterealtime.org/projects/spark/index.jsp">http://www.igniterealtime.org/projects/spark/index.jsp</a>. 17/10/2013

SparkWeb foi desenvolvido utilizando HTML, Javascript e Ajax permitindo que o bate-papo seja acessado via Web.

Para implementar a abordagem proposta no SparkWeb foi identificado o momento em que ele recebe a mensagem que o usuário digitou e que será enviada para o bate-papo. Assim, antes de enviar a mensagem para a tela, um serviço Web que aplicará a abordagem desenvolvida na mensagem enviada é chamado, passando como parâmetros a mensagem, o usuário e o id do bate-papo. Enquanto o serviço Web processa a mensagem, o cliente envia a mensagem para a tela para que os usuários já possam visualizá-la. O serviço Web executa a abordagem proposta, analisando a última mensagem enviada em conjunto com o Comunicografo gerado anteriormente, e retorna um conjunto chave/valor contendo as palavras chaves e os textos das recomendações retornadas para elas.

Finalmente, para apresentação das recomendações na tela, as palavras chaves são apresentadas como *ligações* em uma caixa no canto esquerdo do da interface do sistema (figura 24). Assim, ao passar o mouse em cima dela, um texto com a recomendação é apresentado para o usuário. Dessa maneira, ele pode visualizar apenas as recomendações que interessam a ele, não atrapalhando o fluxo do bate-papo.

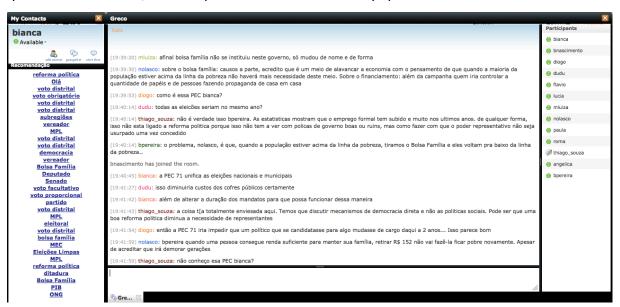


Figura 24: Demonstração das recomendações apresentadas na ferramenta implementada

#### 4.2 Primeira Experimentação

A primeira experimentação foi realizada em sala de aula, com vinte e um alunos de uma turma de graduação em Ciência da Computação na UFRJ em dois dias. No primeiro dia,

uma visão geral da abordagem e os problemas motivadores para essa pesquisa foram abordados. No segundo dia, os alunos participaram de uma sessão de bate-papo na ferramenta desenvolvida em que dois assuntos foram abordados.

Primeiramente, os alunos foram estimulados a dialogar sobre as manifestações ocorridas no Brasil em 2013, um assunto muito atual que gera grandes discussões e diferentes opiniões. O assunto abrange a legislação brasileira e leis existentes com relação a contextos como: manifestações, violência, bloqueio de vias públicas e etc.

Posteriormente, para testar um assunto mais simples em que dúvidas não são comuns, os alunos foram incentivados a discutir sobre turismo, viagens e destinos.

Finalmente, após o diálogo, os alunos responderam o questionário construído com o objetivo de avaliar a abordagem.

Essa experimentação permitiu que fossem analisadas falhas e problemas na abordagem proposta. Assim foi possível analisar as alterações necessárias e implementa-las antes de realizar a segunda experimentação.

No momento em que ela foi realizada as recomendações eram feitas na própria mensagem enviada, marcando uma ou mais palavras e mostrando o texto recomendado quando elas eram focadas com o mouse. Outra diferença foi que a verificação da similaridade semântica entre o texto recomendado e o diálogo ainda não era realizada.

Portanto, com o objetivo de analisar e comparar o desempenho da abordagem nas duas experimentações a avaliação dos questionários respondidos em ambas será discutida na seção 4.4.

#### 4.2.1 Avaliação de Perguntas Abertas

Nessa seção, os resultados das perguntas abertas respondidas no questionário a partir da execução da primeira experimentação serão apresentados.

A primeira questão aberta do questionário pergunta aos participantes quais os problemas, que eles perceberam, que ocorreram durante o bate-papo ou que ocorrem geralmente em bate-papos com grupos de usuários e a abordagem proposta não solucionou ou não solucionaria.

Os participantes comentaram:

"Palavras importantes que deveriam ser grifadas não foram."

O participante percebeu que não foram encontradas recomendações para palavras que ele considera importantes. Analisando as recomendações feitas durante o bate-papo foi possível perceber que o algoritmo de entidades nomeadas não encontrava alguns conceitos. Como solução, um algoritmo de análise morfológica foi adicionado para complementar a recuperação de entidades.

"Alguns termos que encontra, não precisaria recomendar nada. Também acaba sendo muito dependente da riqueza e adequação das fontes para o contexto do Chat."

O participante cita a dependência entre as recomendações e os datasets escolhidos. A base de dados da DBpedia em português contém uma variedade menor de informação do que a sua versão em inglês. De acordo com a wiki<sup>32</sup> da DBpedia a versão em inglês contém 4.004.000 resumos de conceitos, já a versão em português contém apenas 736.000 resumos. Esse fato diminui a riqueza das recomendações e poderia ser solucionado a partir do crescimento da DBpedia PT e do aumento de ligações dessa base com outras bases de dados,. Além disso, também é possível aumentar a quantidade de bases de dados utilizadas pela ferramenta, proporcionando recomendações mais ricas.

"O contexto de uma recomendação ficou incoerente"

Nesse caso o participante encontrou uma recomendação incoerente com o assunto discutido no diálogo. Como contribuição para a solução desse problema, a abordagem passou a usar um algoritmo de similaridade semântica para verificar se as recomendações são estão de acordo com o bate-papo.

A segunda questão aberta pergunta aos participantes quais os problemas, que eles perceberam, que ocorreram durante o bate-papo ou que ocorrem geralmente em bate-papos com grupos de usuários e a abordagem proposta solucionou ou solucionaria.

Os participantes destacaram, principalmente, a recomendação de conceitos pouco conhecidos por eles:

"Informações sobre os termos que estão sendo discutidos que muitas vezes não é bem conhecido pela pessoa."

\_

The DBpedia data set (3.9). Disponível em: <a href="http://wiki.dbpedia.org/Datasets">http://wiki.dbpedia.org/Datasets</a>. 05/01/2013

"Indicação de termos que foram citados e que em parte um usuário não tinha conhecimento."

"Certamente situa muito bem o usuário com as palavras que foram recomendadas."

A terceira e última pergunta aberta solicitava sugestões e comentários adicionais caso necessário. Algumas sugestões interessantes são destacadas a seguir:

"Seria legal que fosse possível responder diretamente uma pessoa."

O participante sugere que seja implementado um mecanismo de encadeamento da conversa para diminuir a perda de co-texto. Já existentes ferramentas de bate-papos, revisadas nessa pesquisa, que utilizam abordagens para isso.

"Caracteres especiais não deveriam ser considerados."

Um participante identificou que ao usar caracteres especiais na mensagem, ocorreu um problema na construção do link com a recomendação. Esse problema foi resolvido para a segunda experimentação.

"Alterou o curso da discussao por escrever na tela onde as mensagens aparecem. A abordagem usando baloes flutuantes me pareceu mais elegante em relacao a esse problema."

O participante identificou que o mecanismo de recomendação implementado, marcando palavras no próprio diálogo, atrapalhava o andamento do bate-papo. Para contribuir para a solução do problema, a partir da segunda experimentação, as recomendações passaram a ser visualizadas em uma caixa de texto no canto esquerdo da tela.

## 4.3 Segunda Experimentação

A segunda experimentação foi realizada a partir da simulação de um bate-papo sobre um dos temas em destaque sendo discutidos na plataforma e-Democracia. Como já descrito, na plataforma os bate-papos ocorrem em comunidades que representam contextos específicos. Devido aos últimos acontecimentos no Brasil, uma das comunidades em destaque na plataforma discute o assunto reforma política<sup>33</sup>.

E-democracia – Reforma Política. Disponível em: <a href="http://edemocracia.camara.gov.br/web/reforma-politica">http://edemocracia.camara.gov.br/web/reforma-politica</a>. 05/11/2013

Portanto, para simular o bate-papo, dezessete participantes que aceitaram discutir o assunto, no bate-papo desenvolvido, participaram da conversa. Para participar do bate-papo apenas é necessário que o participante saiba acessar a Web, a URL do bate-papo, entender o diálogo e interagir com os outros participantes. Portanto, o perfil dos participantes variou entre estudantes de graduação, estudantes de mestrado e doutorado, além de outras pessoas interessadas no assunto de reforma política.

O experimento foi realizado de forma remota na plataforma implementada disponível na Web. Para acessar o bate-papo, os participantes criaram seus usuários em uma URL disponível e entraram na sala de bate-papo previamente criada para a realização do experimento.

Antes do experimento, para que os participantes entendessem o funcionamento da ferramenta e objetivo da mesma, orientações foram disponibilizadas explicitando a motivação para a abordagem proposta e os problemas que ela visa solucionar. Assim, os participantes puderam avaliar as recomendações e a experiência proporcionada ao longo do diálogo.

Finalmente, para avaliar o experimento foi solicitado que os participantes respondessem o questionário construído nessa pesquisa.

#### 4.3.1 Avaliação do Diálogo

A avaliação dos resultados da segunda experimentação descrita nessa seção tem o objetivo de analisar o diálogo gerado pelo bate-papo e o Comunicografo construído pela ferramenta.

A reforma política é um assunto complexo composto por diversas propostas de emenda constitucional, leis e projetos de lei. Por ser um assunto muito extenso foi possível perceber que as pessoas tinham dificuldade no entendimento de diversos termos além da falta de conhecimento das propostas existentes no tema.

Para que o diálogo fosse estimulado, os tópicos que fazem parte do contexto de reforma política foram reunidos e utilizados pelo pesquisador durante o bate-papo para dar continuidade ao diálogo caso o tópico sendo discutido já estivesse esgotado.

Durante o diálogo foi possível perceber diversos momentos em que um participante tinha uma dúvida ou faltava a ele conhecimento para acompanhar o diálogo. Como exemplo, alguns usuários não conheciam o texto das propostas de emenda a constituição (figura 25).

```
(19:41:59) thiago_souza: não conheço esa PEC bianca?
```

Figura 25: Trecho do diálogo realizado na experimentação

Na figura 26 pode-se verificar um exemplo em que a recomendação contribuiu para solucionar a dúvida da participante. A proposta de voto distrital não era conhecida por ela e foi recomendada pela ferramenta.

```
bianca: como o alan deu a ideia podemos começar pela proposta de voto distrital, o que vocês acham bpereira: legal, eu não conhecia a proposta e a recomendação me ajudou
```

Figura 26: Trecho do diálogo realizado na experimentação

Logo adiante, conforme evidenciado na pergunta da figura 27, o participante demonstra que também não tem conhecimento sobre o assunto.

```
thiago: como funcionaria esse tipo de voto? Seria dividido por estados?
```

Figura 27: Trecho do diálogo realizado na experimentação

Assim como a participante *usuario2*, a dúvida do participante também poderia ser solucionada pela mesma recomendação.

Também foi possível perceber que o bate-papo com essa quantidade de usuários fica confuso, muitos usuários digitam ao mesmo tempo e às vezes acontece sobreposição de assuntos.

```
(20:09:42) thiago_souza: mluiza, verdade, mas não estamos tão mal assim . Junho mostrou que os anos de articulação dos movimentos sociais deram em alguma coisa. Tanto que estivemos 6 meses na rua
(20:09:53) nolasco: mas quem não quer entender é só não votar!
beatriz has joined the room.
(20:10:08) bianca: concordo nolasco
(20:10:08) mluiza: hum, mesmo? Nao vi continuidade nem propostas...
(20:10:12) angelica: Concordo, Luiza. Voltamos a educação política...
(20:10:17) thiago_souza: concordo com a angelica. Pelbiscito é pra coisas pontuais
(20:10:58) alfredo: e temos massa crítica politizada o suficiente para mobilizações de monta? acho que não...
(20:11:14) angelica: E educação política se dá nas escolas, nas residências ... Como fazer?
(20:11:14) nolasco: thiago como você acha que um sistema por voto majoritária irá satisfazer as minorias?
(20:11:16) bianca: temos algumas propostas, foi criado um grupo de trabalho da Câmara dos Deputados que está definindo algumas PECs e projetos de lei
(20:11:19) thiago_souza: alfredo, como assim?
```

Figura 28: Trecho do diálogo realizado na experimentação

Nesse caso, na figura 28, é possível perceber que vários assuntos estão sendo discutidos ao mesmo tempo e o usuário *thiago\_souza* não consegue acompanhar o diálogo.

Nesse trecho quatro assuntos estavam sendo discutidos ao mesmo tempo. Alguns participantes estavam discutindo sobre o plebiscito para a reforma política quando dois participantes começaram a conversar sobre o problema do desconhecimento de política por parte da população. Simultaneamente, dois outros participantes falam sobre manifestações e finalmente a última mensagem, desconexa das anteriores, fala sobre propostas para reforma.

Nessa pesquisa, a visualização do Comunicografo não foi disponibilizada para os participantes. Porém, ela foi gerada para analisar o Comunicografo gerado no experimento. Visualizando o Comunicografo completo da discussão é possível perceber que essa mensagem respondia a uma pergunta feita quinze mensagens e vinte e três segundos antes. Esse fato mostra a complexidade do bate-papo com muitos participantes pois caso um usuário demore a responder uma pergunta, a resposta pode parecer desconexa para outros participantes.

Portanto, a partir dessa análise é possível perceber que a visualização no formato de grafo poderia ser disponibilizada para os participantes ao longo do diálogo. Assim, seria fácil perceber as mensagens que estão relacionadas, visualizando os assuntos, e entender melhor a estrutura do diálogo, possibilitando também a solução de problemas de perda de co-texto. Na figura 29 os assuntos discutidos nesse trecho são destacados.

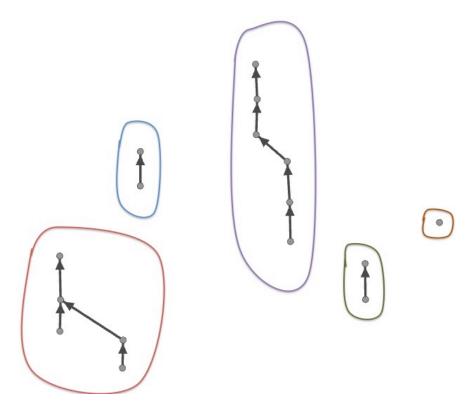


Figura 29: Referências entre mensagens no trecho do diálogo realizado na experimentação

O bate-papo realizado na experimentação gerou duzentas e treze mensagens que
foram encadeadas no Comunicografo armazenado no formato XML.

## 4.3.2 Avaliação de Perguntas Abertas

O questionário respondido pelos participantes é composto de questões obrigatórias fechadas que apresentam como resposta uma múltipla escolha baseada na escala de Likert e questões abertas opcionais que tem o objetivo de obter opiniões, sugestões e criticas dos participantes.

Inicialmente, as questões abertas e suas respostas serão discutidas. Posteriormente, na seção 4.4 será feita a análise das questões obrigatórias respondidas em conjunto com as questões respondidas na primeira experimentação.

A primeira questão aberta pergunta aos participantes quais os problemas percebidos que ocorreram durante o bate-papo ou que ocorrem geralmente em bate-papos com grupos de usuários e a abordagem proposta não solucionou ou não solucionaria.

Alguns participantes destacaram:

"As vezes recomenda ainda termos bastante triviais"

"Talvez alguns termos muito comuns pudessem ser eliminados de alguma forma. Também poderia ter um local onde se colocassem termos a recomendar"

"Acho que é preciso de alguma forma eliminar recomendações para termos pouco relevantes. Acho difícil, mas pode introduzir um ruído no chat a recomendação de definições de termos muito óbvios."

"Um termo muito citado, deve ser recomendado? ou apenas algo que realmente exista a necessidade de desambiguação ou necessidade de obtenção de conhecimento sobre o termo?"

Inicialmente, o primeiro participante destaca que alguns termos desnecessários foram recomendados. Outros participantes, posteriormente, também citaram esse problema. A análise desse fato e as propostas de solução são abordadas na seção 3.4.

O primeiro participante cita a dependência entre as recomendações e os datasets escolhidos. A base de dados da DBpedia em português contém uma variedade menor de informação do que a sua versão em inglês. De acordo com a wiki<sup>34</sup> da DBpedia a versão em inglês contém 4.004.000 resumos de conceitos, já a versão em português contém apenas 736.000 resumos. Esse fato diminui a riqueza das recomendações e poderia ser solucionado a partir do crescimento da DBpedia PT e do aumento de ligações dessa base com outras bases de dados, proporcionando recomendações mais ricas.

Outro participante comentou:

"Algumas palavras tem descrição errada, levando ao conhecimento errôneo de certas pessoas"

Esse problema depende dos datasets utilizados e informações disponíveis.

Muitos participantes comentaram sobre as dificuldades encontradas em bate-papos com grupos de usuários como a quantidade de assuntos discutidos ao mesmo tempo e a dificuldade de acompanhar a conversa:

"Dificuldade de acompanhar a conversa devido a grande quantidade de usuários."

\_

The DBpedia data set (3.9). Disponível em: <a href="http://wiki.dbpedia.org/Datasets">http://wiki.dbpedia.org/Datasets</a>. 05/01/2013

"Múltiplos assuntos ao mesmo tempo. A conversa não segue uma linha de raciocínio."

"A perda de foco em assuntos, sem esgotá-los."

"Perda de contexto ao longo das conversas."

"Com excessivos usuários, é fácil de se proliferar segmentos de diferentes assuntos. "

"O desfoque do assunto não está ligado somente ao tema majoritariamente discutido, mas na intenção do usuário em si. Caso o usuário queira mudar o assunto bruscamente, não será uma recomendação que o fará a pensar o contrário. "

A perda de co-texto (PIMENTEL 2002), discutida na revisão de trabalhos relacionados, é um problema que não é solucionado por recomendações de conteúdo pois o participante não consegue acompanhar o encadeamento do diálogo por outros motivos, além da falta de conhecimento. Portanto, seria interessante utilizar mecanismos, descritos em trabalhos relacionados, que permitam ao participante encadear o diálogo explicitamente. Essa abordagem pode complementar as recomendações solucionando esse tipo de problema e facilitando a construção do Comunicografo. Este aspecto será explorado em versões futuras do AidChat.

Outro participante destacou um problema interessante:

"A ferramenta não contribui para exibir os assuntos discutidos durante o chat de forma diferenciada. Como diversos assuntos foram discutidos, seria interessante a ferramenta exibir esses assuntos de forma diferenciada, para facilitar o entendimento pelos participantes do chat."

Como a quantidade de assuntos simultâneos é muito grande nesse tipo de bate-papo , sendo muito difícil acompanhar a conversa, o participante sugeriu que a abordagem exibisse os assuntos discutidos separadamente. Essa abordagem pode ser construída recuperando threads de assuntos a partir do Comunicografo, identificando o assunto correspondente e apresentando a visualização do grafo para os participantes. Esta funcionalidade já está implementada, porém não foi disponibilizada para visualização dos participantes.

A segunda questão aberta pergunta aos participantes quais os problemas percebidos que ocorreram durante o bate-papo ou que ocorrem geralmente em bate-papos com grupos de usuários e a abordagem proposta solucionou ou solucionaria.

Alguns participantes destacaram que a ferramenta solucionou algumas dúvidas:

"Conceitos que não são conhecidos pelos usuários puderam ser explanados durante a conversação sem a necessidade de efetuar uma busca, contribuindo para que o usuário não ficasse perdido em relação ao que estava sendo dito "

"Algumas recomendações foram bem uteis para entendimento de alguma referencia feita por algum participante do chat "

"Acho que as principais contribuições da ferramenta é a ajuda ao usuário a se interar do assunto principal que está sendo debatido no chat, fazendo com que ele fique menos perdido, especialmente quando ele acaba de entrar no chat. Outra contribuição importante é a desambiguação das palavras, a explicação que é feita a partir das palavras recomendadas ajuda bastante o usuário a se iterar melhor sobre o assunto e ter a compreensão certa sobre o que está sendo discutido, melhorando a qualidade do debate no chat. "

"A ferramenta ajuda a esclarecer alguns conceitos comentados durante o chat, contribuindo para um melhor entendimento do assunto discutido. Os termos são exibidos e recomendados pela própria ferramenta, o que diminui problemas de entendimento e ambiguidades, garantindo que todas as pessoas estão conversando sobre o mesmo termo."

"Ajudou em dar um contexto sobre o que estava sendo conversado quando eu desconhecia totalmente o assunto e os termos usados. "

Devido à velocidade do bate-papo com vários usuários é necessário que a recomendação seja muito rápida para solucionar as dúvidas dos participantes. Porém, devido às consultas feitas em bases online e servicos Web, e ao processamento intenso, a recomendação demora a aparecer. Portanto, é necessário avaliar métodos para melhorar o desempenho da abordagem implementada.

A terceira e última pergunta aberta solicitava sugestões e comentários adicionais caso necessário. Algumas sugestões interessantes são destacadas a seguir:

"Algumas recomendações aparecem multiplicadas. "

"Sim, notei haver termos repetidos nas recomendações. "

Alguns participantes destacaram que recomendações aparecem repetidas. O algoritmo não verifica se uma recomendação já existe na tela para que um novo participante também possa visualizar a recomendação caso a entidade seja citada. Porém, seria mais interessante que, caso um novo usuário entrasse, as recomendações anteriores fossem carregadas na tela.

"Algumas definições poderiam ser melhores mesmo que o texto ficasse um pouco maior."

O participante destacou que os resumos que descrevem as entidades poderiam ser maiores. Porém, o resumo depende da informação armazenada das bases de dados escolhidas, nesse caso a DBpedia e o LexML, além de ser importante que o resumo seja sucinto para que a visualização seja rápida.

## 4.4 Avaliação de perguntas Fechadas

As questões fechadas foram divididas em dois grupos no questionário. Foram definidas quatro perguntas relativas a bate-papos em geral e cinco perguntas relativas à ferramenta proposta. Os resultados das respostas dadas pelos participantes para questões fechadas do questionário foram analisados a partir de gráficos que representam a porcentagem de participantes que marcou cada resposta em cada uma das experimentações.

#### Perguntas relativas a bate-papos em geral

1. Você já teve ou presenciou dificuldades de comunicação em bate-papos (Devido a palavras ambíguas e falta de entendimento ou conhecimento sobre o assunto).

As opções disponíveis eram: Discordo totalmente (1), Discordo em grande parte (2), Discordo em parte (3), Nem concordo nem discordo (4), Concordo em parte (5), Concordo em grande parte (6), Concordo totalmente (7)

#### Primeira Experimentação

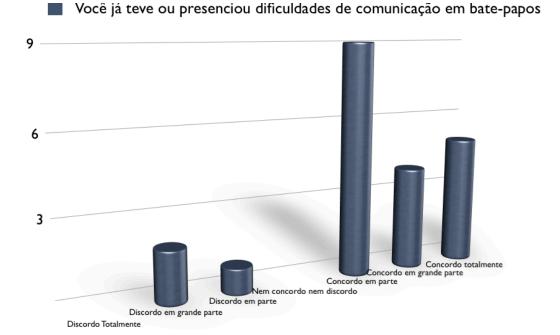


Figura 30: Análise da quantidade de participantes que marcaram cada resposta

## Segunda Experimentação

Você já teve ou presenciou dificuldades de comunicação em bate-papos

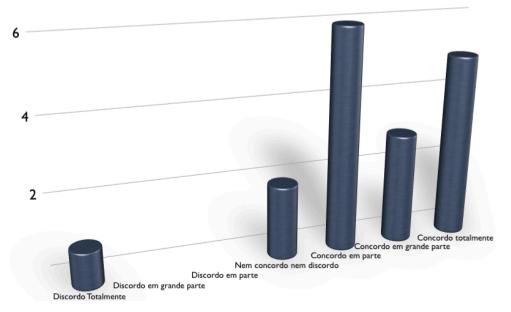


Figura 31: Análise da quantidade de participantes que marcaram cada resposta

## Considerações

As respostas marcadas nas duas experimentações indicam que a maioria dos participantes já teve experiências com problemas de comunicação em ambientes de comunicação textual. Isso destaca a necessidade de contribuições que diminuam os diversos problemas que podem acontecer no demorrer de um diálogo realizado em um bate-papo.

#### 2. Em chats com múltiplos usuários é difícil de acompanhar a conversa.

As opções disponíveis eram: Discordo totalmente (1), Discordo em grande parte (2), Discordo em parte (3), Nem concordo nem discordo (4), Concordo em parte (5), Concordo em grande parte (6), Concordo totalmente (7)

## Primeira Experimentação

Em chats com múltiplos usuários é difícil de acompanhar a conversa.

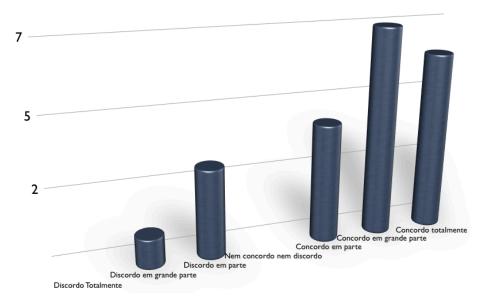


Figura 32: Análise da quantidade de participantes que marcaram cada resposta

## Segunda Experimentação

Em chats com múltiplos usuários é difícil de acompanhar a conversa.

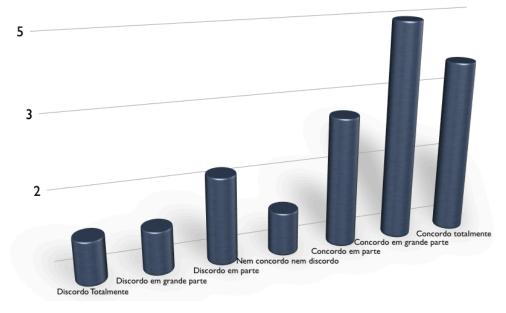


Figura 33: Análise da quantidade de participantes que marcaram cada resposta

#### Considerações

Portanto, nas figuras 34 e 35 é possível visualizar que os participantes acreditam que em bate-papos com grupos de usuários é difícil acompanhar a conversa. O objetivo dessa pergunta foi perceber se os participantes concordam que em bate-papos com grupos de usuários a dificuldade de acompanhar a conversa é maior, tornando necessário o desenvolvimento de contribuições que apoiem a diminuição desse problema.

## 3. Qual o nível de dificuldade que você percebe para fazer uma pergunta sobre o assunto da conversa em um bate-papo com múltiplos usuários?

As opções disponíveis eram: Extremamente fácil (1), Muito fácil (2), Fácil (3), Nem fácil nem difícil (4), Difícil (5), Muito Difícil (6), Extremamente difícil (7)

#### Primeira Experimentação

Qual o nível de dificuldade que você percebe para fazer uma pergunta sobre o assunto da conversa em um bate-papo com múltiplos usuários?

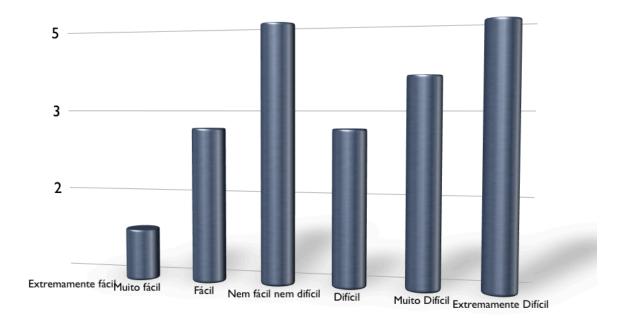


Figura 34: Análise da quantidade de participantes que marcaram cada resposta Segunda Experimentação

2

📕 Qual o nível de dificuldade que você percebe para fazer uma pergunta sobre o assunto da conversa em um bate-papo com múltiplos usuários?

Figura 35: Análise da quantidade de participantes que marcaram cada resposta

Muito Difícil Extremamente Difícil

Nem fácil nem difícil

Fácil

#### Considerações

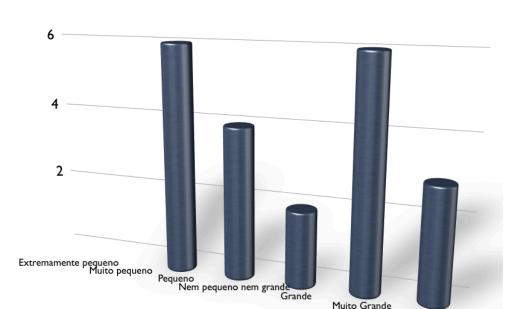
Extremamente fáci Muito fácil

O objetivo da pergunta respondida nas figuras 36 e 37 era perceber se os participantes tem dificuldade para fazer uma pergunta, ao longo do diálogo, sobre um conceito que eles não entenderam ou não tem conhecimento. Porém, talvez, os participantes não tenham entendido muito bem a pergunta. Na primeira experimentação, apesar da maioria ter marcado que é difícil, muitos participantes responderam que é fácil ou nem fácil nem difícil. Na segunda experimentação a maioria das pessoas tendeu a responder que é fácil. Porém, na experimentação realizada percebeu-se que quando um participante tenta tirar uma dúvida sobre um conceito que não entendeu, devido a velocidade do diálogo, muitas vezes ele não é respondido, ou a resposta é demorada fazendo com que ele não consiga mais acompanhar o diálogo.

# 4. Qual o impacto, sobre o diálogo, de fazer uma pergunta, para sanar uma dúvida, durante o bate-papo?

As opções disponíveis eram: Extremamente pequeno (1), Muito pequeno (2), Pequeno (3), Nem pequeno nem grande (4), Grande (5), Muito grande (6), Extremamente grande (7)

Primeira Experimentação



Qual o impacto, sobre o diálogo, de fazer uma pergunta, para sanar uma dúvida, durante o bate-papo

Figura 36: Análise da quantidade de participantes que marcaram cada resposta

Extremamente Grande

## Segunda Experimentação

Qual o impacto, sobre o diálogo, de fazer uma pergunta, para sanar uma dúvida, durante o bate-papo

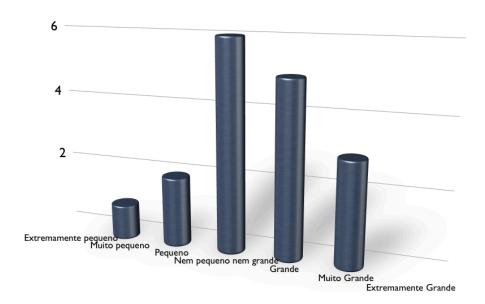


Figura 37: Análise da quantidade de participantes que marcaram cada resposta

## Considerações

Talvez essa pergunta tenha sido mais clara que a anterior. O objetivo é entender se os participantes acham que a tentativa de sanar uma dúvida durante o diálogo pode atrapalhar o andamento da discussão. Nesse caso, na figura 37, os participantes ficaram divididos. Muitos responderam que o impacto é pequeno ou grande. Na figura 38 percebe-

se que a maioria dos participantes esteve em dúvida. O resultado indica que muitos deles não percebem o impacto causado ao fazer uma pergunta que atrapalha o diálogo diminuindo o foco no assunto.

### Perguntas com relação a ferramenta proposta

#### 1. Suas dúvidas com relação ao assunto foram sanadas a partir das recomendações feitas.

As opções disponíveis eram: Discordo totalmente (1), Discordo em grande parte (2), Discordo em parte (3), Nem concordo nem discordo (4), Concordo em parte (5), Concordo em grande parte (6), Concordo totalmente (7)

## Primeira Experimentação

Suas dúvidas com relação ao assunto foram sanadas a partir das recomendações feitas.

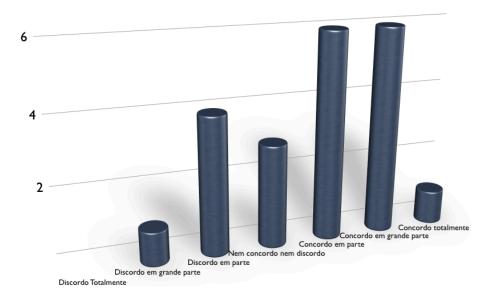
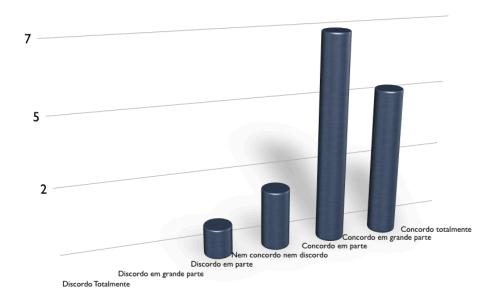


Figura 38: Análise da quantidade de participantes que marcaram cada resposta

#### Segunda Experimentação



Suas dúvidas com relação ao assunto foram sanadas a partir das recomendações feitas.

Figura 39: Análise da quantidade de participantes que marcaram cada resposta

#### Considerações

Na primeira experimentação alguns participantes discordaram da afirmação indicando que as dúvidas não foram sanadas pela ferramenta. Já na segunda um número maior de pessoas afirmou que as recomendações contribuíram com a solução de dúvidas. Isso aconteceu, pois na segunda experimentação foram implementadas melhorias que tornaram o sistema de recomendação mais eficiente. Porém, devido às recomendações apresentadas ainda percebe-se, visualmente, que poucas dúvidas foram solucionadas.

#### 2. As recomendações feitas pela ferramenta foram construtivas para a conversa.

As opções disponíveis eram: Discordo totalmente (1), Discordo em grande parte (2), Discordo em parte (3), Nem concordo nem discordo (4), Concordo em parte (5), Concordo em grande parte (6), Concordo totalmente (7)

#### Primeira Experimentação

As recomendações feitas pela ferramenta foram construtivas para a conversa

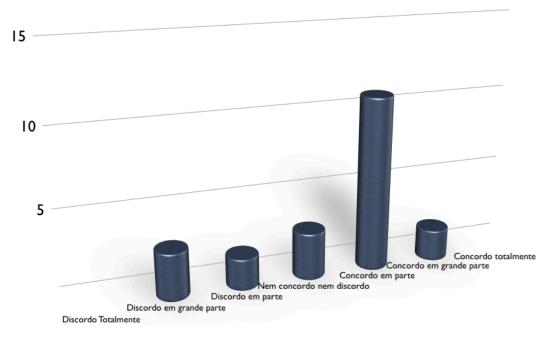


Figura 40: Análise da quantidade de participantes que marcaram cada resposta **Segunda Experimentação** 

As recomendações feitas pela ferramenta foram construtivas para a conversa

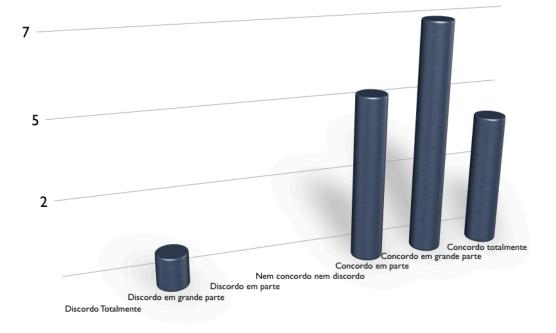


Figura 41: Análise da quantidade de participantes que marcaram cada resposta Considerações

Os resultados dessa questão também indicam que na segunda experimentação um número maior de participantes achou que as recomendações foram construtivas para o diálogo. Na segunda experimentação a maioria dos participantes marcou que as recomendações foram construitivas.

3. Você percebeu que a ferramenta solucionou problemas que geralmente ocorrem em chats com grupos de usuários. (Confusão, perguntas, usuários perdidos, ambiguidades)

As opções disponíveis eram: Discordo totalmente (1), Discordo em grande parte (2),

Discordo em parte (3), Nem concordo nem discordo (4), Concordo em parte (5), Concordo em grande parte (6), Concordo totalmente (7)

## Primeira Experimentação

Você percebeu que a ferramenta solucionou problemas que geralmente ocorrem em chats com multiplos usuários.

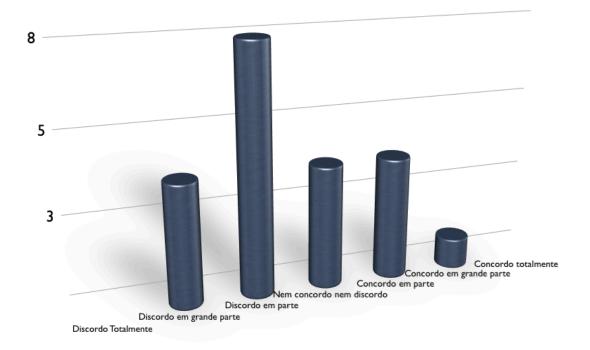


Figura 42: Análise da quantidade de participantes que marcaram cada resposta Segunda Experimentação

2

Nem concordo em parte

Discordo em grande parte

Discordo em grande parte

Você percebeu que a ferramenta solucionou problemas que geralmente ocorrem em chats com multiplos usuários.

Discordo Totalmente

Figura 43: Análise da quantidade de participantes que marcaram cada resposta

## Considerações

Na avaliação feita na primeira experimentação grande parte dos participantes discordou da afirmação, mostrando que a abordagem não solucionou os diversos problemas que podem ocorrer em um bate-papo. Um exemplo é a perda de co-texto em que o usuário não consegue acompanhar a discussão devido a outros motivos, diferentes de falta de conhecimento.

# 4. Qual a frequência(na sua opinião) que um usuário comum teria que fazer buscas na Web, no cenário utilizado, caso não fossem feitas recomendações?

As opções disponíveis eram: Extremamente baixa (1), Muito baixa (2), Baixa (3), Nem baixa nem alta (4), Alta (5), Muito alta (6), Extremamente alta (7)

#### Primeira Experimentação

Qual a frequência(na sua opinião) que um usuário comum teria que fazer buscas na web, no cenário utilizado, caso não fossem feitas recomendações?

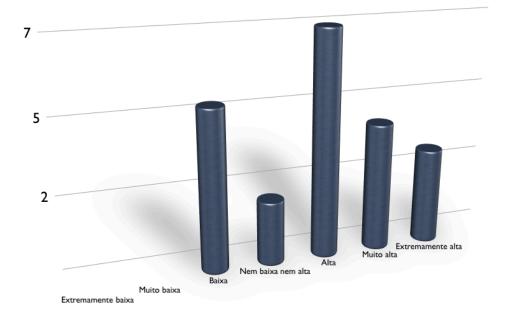


Figura 44: Análise da quantidade de participantes que marcaram cada resposta

## Segunda Experimentação

Qual a frequência(na sua opinião) que um usuário comum teria que fazer buscas na web, no cenário utilizado, caso não fossem feitas recomendações?

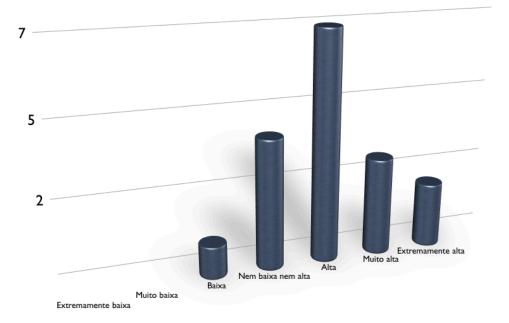


Figura 45: Análise da quantidade de participantes que marcaram cada resposta **Considerações** 

Essa questão visa identificar se os participantes acham que no caso de uma ferramenta comum de bate-papo, em que recomendações não são feitas, seria necessário fazer buscas na Web para encontrar dados sobre conceitos que eles desconhecem. Na primeira experimentação a maioria dos participantes respondeu que a frequência seria alta, porém uma grande quantidade de pessoas também respondeu que a frequência seria baixa, provavelmente porque essa experimentação discutiu assuntos mais simples. Na segunda experimentação em que um assunto muito complexo foi discutido a maioria das pessoas responderam que a frequência seria alta.

## 5. A ferramenta recomendou informações desnecessárias. (Já conhecidas por grande parte da população ou desnecessárias para o contexto do diálogo)

As opções disponíveis eram: Discordo totalmente (1), Discordo em grande parte (2), Discordo em parte (3), Nem concordo nem discordo (4), Concordo em parte (5), Concordo em grande parte (6), Concordo totalmente (7)

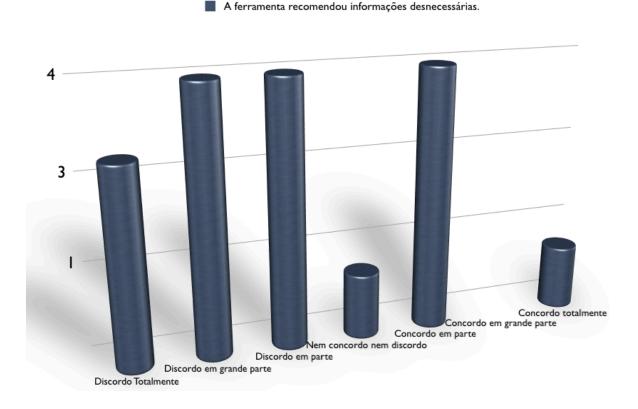


Figura 46: Análise da quantidade de participantes que marcaram cada resposta Considerações

A quinta questão só foi adicionada na segunda experimentação. Ela tem o objetivo de verificar um problema percebido na primeira experimentação, a recomendação de conceitos desnecessários. O resultado da questão mostra que a maioria das pessoas não concordou com a afirmação e discordou do fato de que recomendações desnecessárias foram feitas. Porém, também se pode perceber que muitos participantes concordaram que algumas recomendações foram desnecessárias.

## 5 Conclusão

O aumento da utilização da Internet, bem como dos meios de acesso a ela, além da popularização da Web 2.0 provocaram um considerável aumento na utilização dos meios de comunicação textuais. Esse trabalho se apoia nesse cenário, como motivação, para propor uma abordagem que contribua com a experiência de participação de um bate-papo online.

Isto contribuiu para a percepção das facilidades que os meios de comunicação textuais podem trazer para o usuário da Web. Assim, atualmente, é comum que estes meios sejam usados diariamente para a troca de e-mails, conversas rápidas em redes sociais e troca de mensagens a partir de dispositivos móveis. Porém, além do aumento da utilização desses meios para tarefas simples e corriqueiras, eles também passaram a ser muito usados em cenários corporativos e governamentais, e, por consequência, mais complexos. Alguns exemplos são ferramentas de mensagens instantâneas para o ambiente corporativo, ferramentas de suporte a usuários e ferramentas de comunicação textual para vendas em sistemas de comércio eletrônico.

Essa pesquisa se concentra nestas utilizações mais complexas e nos diálogos que ocorrem entre grupos de usuários, devido à quantidade maior de problemas encontrados e ao grande impacto que estas podem causar. Nesses diálogos, dependendo do nível de conhecimento do participante sobre o assunto e da complexidade do mesmo, é comum que surjam muitas dúvidas e seja difícil acompanhar a conversa.

Portanto, a abordagem AidChat proposta e implementada nesse trabalho tem o objetivo de contribuir com a qualidade de diálogos realizados em ferramentas de bate-papo online, fazendo a análise do bate-papo e disponibilizando recomendações de conteúdo para os participantes. Estas fazem com que eles possam aprender novos conceitos rapidamente na própria ferramenta, diminuindo problemas de entendimento e melhorando a qualidade e resultados do diálogo.

A recomendação de informação complementar em bate-papos, quando aplicada em contextos complexos, pode contribuir para o diálogo, permitindo que os participantes tenham uma base maior para discutir e argumentar. Assim, a discussão torna-se mais construtiva. Como exemplo, no cenário aplicado nessa pesquisa os bate-papos são

realizados ao longo de assembleias públicas tendo o objetivo de que os participantes opinem na criação de projetos de leis e emendas constitucionais. Portanto, o conhecimento do participante sobre o assunto é muito importante para que ele possa colaborar com as decisões tomadas pelos deputados.

O portal e-Democracia foi escolhido como cenário para que o AidChat fosse testado em um ambiente em que dúvidas são muito comuns e os assuntos discutidos são muito complexos. Portanto, a abordagem foi aplicada em ambiente similar à ferramenta de batepapo do portal e-Democracia e como experimentação seções de bate-papo foram realizadas para discutir assuntos que estão em destaque no portal.

Dessa maneira, foi possível também encontrar limitações e sugerir trabalhos futuros que são apresentados nas seções posteriores.

## 5.1 Contribuições Principais

Como principais contribuições dessa pesquisa, pode-se destacar:

- a discussão dos detalhes para a implementação de um Comunicografo automático complementando trabalhos anteriores. Além disso, a possibilidade de obter o grafo automaticamente, durante o diálogo, trouxe como contribuição uma abordagem que pode ser explorada para que, a partir da estrutura do diálogo, sejam realizadas diversas análises e extrações de informações, automáticas.
- a proposta de uma abordagem de recomendação de conteúdo em bate-papos.
   Dentre as abordagens existentes para a recomendação em bate-papos na literatura,
   não existe nenhuma que recomende informações sucintas a partir de dados extraídos do bate-papo. Outra facilidade é permitir que o usuário pudesse enchergar as recomendações na própria tela do bate-papo, tornando a sua visualização muito rápida.
- a utilização da metodologia de LOD para recomendação, permitindo que sejam utilizadas bases de dados referentes a diversos contextos. Assim, apesar da abordagem ter sido aplicada sobre um contexto específico, as bases de dados utilizadas para recomendação podem ser alteradas facilmente, permitindo que ela seja aplicada facilmente em outros contextos.

### 5.2 Limitações

Ao longo da implementação e experimentação da abordagem AidChat diversas limitações foram encontradas. Nesta seção as limitações são discutidas para que trabalhos futuros sejam propostos na próxima seção.

Uma das grandes dificuldades encontradas durante a experimentação foi a velocidade com que a ferramenta apresenta as recomendações. Esse problema faz com que, as vezes, a recomendação só seja visualizada quando o assunto já mudou, fazendo com que o usuário perca a chance de argumentação e entendimento do diálogo. Além disso, a representação da comunicação é simplificada devido a dificuldade computacional, permitindo que cada mensagem só possa fazer referência a uma outra mensagem. Essa simplicação implica na representação da referência entre as mensagens mais próximas, perdendo informação. Um exemplo é quando uma mensagem discute ou responde a diversas mensagens anteriores.

Outro problema é que algumas entidades recomendadas são pouco relevantes para o contexto sendo discutido. Essas entidades representam conceitos muito comuns, amplamente conhecidos, que não proporcionam aprendizado e atrapalham a visualização de outras entidades mais construtivas para o diálogo.

Outra dificuldade, destacada por um participante da experimentação, é a dependência de bases de dados escolhidas para a recomendação. Portanto, dependendo das fontes utilizadas e do contexto do diálogo o dempenho durante a recomendação pode variar muito.

Além disso, outro problema percebido foi a repetição de conceitos recomendados que pode confundir os usuários e atrapalhar a visualização de outros conceitos.

A abordagem de inferência de mensagens referenciadas por dúvida apresenta uma necessidade de aprimoramento. Nessa pesquisa, quando as dúvidas são curtas e não fazem referências a outras mensagens não é possível identificar o conceito ou assunto ao qual elas se referem.

### **5.3 Trabalhos Futuros**

Finalmente, a partir da análise das limitações, serão discutidas abordagens que podem diminuir o impacto causado por elas, além de outras evoluções interessantes.

Na abordagem proposta, a recomendação de conteúdo permitiu que algumas dúvidas fossem solucionadas na própria ferramenta. Porém, devido à quantidade de análises feitas e consultas executadas para que a recomendação fosse recuperada, as informações complementares demoram a aparecer na tela. Portanto, como trabalho futuro é necessário que os algoritmos sejam aprimorados para que a recomendação apareça mais rápida para os participantes.

Outra melhoria que deveria ser analisada seria permitir ao usuário navegar na recomendação a partir de Web de Dados bem como suas ligações, fazendo com que seu aprendizado fosse ainda maior. Assim o usuário poderia visualizar melhor o conceito e todos os seus atributos. Porém, seria necessário que essa navegação ocorresse sem sair do ambiente da ferramenta e fosse sucinta, pois em com mais de dois usuários a velocidade com que as mensagens são enviadas é muito grande e o usuário não conseguiria mais acompanhar caso ele tivesse que sair da ferramenta. Outro problema que deve ser analisado é o nível de atenção dos participantes, pois, talvez, a partir da possibilidade de navegação ele demorasse a voltar a ter atenção no diálogo.

Outra ideia para trabalho futuro, em consequência da limitação de recomendações repetidas, seria utilizar o conceito de *word clouds*, em que as palavras que são chaves das recomendações seriam reunidas em grupos de palavras que aumentam de acordo com o número de vezes em que a palavra aparece no diálogo. Porém, seria necessário que as palavras recomendadas naquele diálogo fossem carregadas sempre que um novo usuário entrasse no bate-papo, pois ele pode não conhecer algum conceito apresentado anteriormente.

Futuramente, algoritmos de recomendação de dados abertos podem ser implementados de maneira complementar ao algoritmo desenvolvido. Assim, seria possível encontrar entidades similares a entidades mencionadas pelos usuários. No exemplo abaixo, retirado do bate-papo realizado na experimentação, o usuário menciona o projeto ficha limpa<sup>35</sup>.

"."

Figura 49: Exemplo de mensagem enviada durante bate-papo realizado na ferramenta AidChat

\_

<sup>35</sup> Ficha Limpa. Disponível em: <a href="http://dbpedia.org/page/Ficha\_Limpa">http://dbpedia.org/page/Ficha\_Limpa</a>. 27/12/2013

Assim, o algoritmo poderia identificar que o recurso "Consolidação das leis do trabalho"<sup>36</sup> é similar à lei da ficha limpa pois também é uma lei brasileira descrita pelo recurso "Brazilian Law"<sup>37</sup>. Porém, seria necessário aplicar um meio de avaliação das recomendações de recursos similares para identificar se são realmente informações relevantes.

Outra melhoria muito interessante, possível a partir da utilização de conceitos de LOD é a publicação de dados extraídos do diálogo. Assim, poderia ser criada uma base em RDF, que seria interligada com outras bases disponíveis na WoD, e publicaria conceitos recomendados categorizados por contextos que seriam extraídos do diálogo. Outra possibilidade é a publicação de dados dos participantes, como interesses conhecimentos e assuntos em que ele tem pouco domínio. Assim, essas informações poderiam ser reutilizadas em outros diálogos e até ferramentas.

Além disso, também é necessário que a inferência dos conceitos referenciados por dúvidas seja aprimorada. Sendo ela o principal problema que a pesquisa tem o objetivo de contribuir para a solução, é interessante que, além de recuperar as mensagens referenciadas por ela, existam outros meios para identificar a que assunto ela se refere. Talvez, dúvidas representadas por mensagens pequenas, em que não é possível encontrar uma referência, como: "não entendi" ou "pode repetir?", façam referência a última mensagem enviada devido a rapidez com que elas podem ser digitadas. Além disso, também poderia ser mantida uma lista de ações que serão realizadas de acordo com a palavra utilizada para indicar uma dúvida.

De acordo com a experiência obtida nas experimentações foi possível perceber que a velocidade na qual o usuário visualiza a recomendação e a interface em que ela é disponibilizada podem melhorar a contribuição da aplicação. Portanto, é necessário identificar os diversos perfis de participantes e identificar a melhor maneira de apresentação dessa informação de maneira rápida e compreensível.

<a href="http://dbpedia.org/page/Consolidation\_of\_Labor\_Laws">http://dbpedia.org/page/Consolidation\_of\_Labor\_Laws</a>. 27/12/2013

-

Consolidação das leis do trabalho. Disponível em:

Portanto, esse trabalho apresentou uma abordagem para enriquecer a experiência de bate-papo online identificando diversas limitações e contribuições. Assim foi possível propor diversos trabalhos futuros que podem evoluir e enriquecer a área de análise da comunicação.

### 6 Referências

BAKHTIN, M. Problemas da poética de Dostoiévski. Forense Universitária, 1981

BASTIAN, M.; HEYMANN, S.; JACOMY, M. **Gephi: an open source software for exploring and manipulating networks**. International AAAI Conference on Weblogs and Social Media. 2009.

BIZER, C. **The Emerging Web of Linked Data**, Intelligent Systems, IEEE , vol.24, no.5, pp.87,92, 2009

CHANDRA, P.; CAMBRIA, E.; PRADEEP, A. Enriching social communication through semantics and sentics. Sentiment Analysis where AI meets Psychology (SAAIP), p. 68, 2011.

DAIBER, J. et al. Improving Efficiency and Accuracy in Multilingual Entity Extraction. Proceedings of the 9th International Conference on Semantic Systems (I-SEMANTICS '13). ACM, New York, NY, USA, 121-124.

DE PAIVA, V.; COMMERCE, R.; RADEMAKER, A. **Revisiting a Brazilian WordNet**. In: GWC 2012 6th International Global Wordnet Conference. 2012. p. 100.

DOERR, Martin et al. **Frbroo, a conceptual model for performing arts**. 2008 Annual Conference of CIDOC, Athens, September 15 – 18, 2008.

DOMAVICIUS, G.; TUZHILIN, A. Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. Knowledge and Data Engineering, IEEE Transactions on, v. 17, n. 6, p. 734-749, 2005.

FUKS, H. **Groupware technologies for education in AulaNet**. Computer Applications in Engineering Education, v. 8, n. 3-4, p. 170-177, 2000.

FUKS, H.; GEROSA, M. A.; LUCENA, C. J. P. de. **Usando a categorização e estruturação de mensagens textuais em cursos pelo ambiente AulaNet**. Revista Brasileira de Informática na Educação, v. 10, n. 1, p. 31-44, 2002.

FUKS, H.; PIMENTEL, M. **Studying Response-Structure Confusion in VMT**. In: G.G. Stahl. (Org.). Virtual Math Teams. : Springer Science+Business Media, 2009, p. 373-398.

GABRILOVICH, E.; MARKOVITCH, S. Computing Semantic Relatedness Using Wikipedia-based Explicit Semantic Analysis. In Proceedings of the 20th International Joint Conference on Artificial Intelligence. 2007. p. 1606-1611.

GARIJO, D.; VILLAZÓN-TERRAZAS, B.; CORCHO, O. **A provenance-aware linked data application for trip management and organization**. In: Proceedings of the 7th International Conference on Semantic Systems. ACM, 2011. p. 224-226.

GERMESIN, S.; WILSON, T. **Agreement detection in multiparty conversation**. In: Proceedings of the 2009 international conference on Multimodal interfaces. ACM, 2009. p. 7-14.

GHAHRAMANI, Z. **Learning dynamics Bayesian networks.** Lecture Notes in Computer Science Volume 1387, 1998, pp 168-197

HARTIG, O.; BIZER, C.; FREYTAG, J. **Executing SPARQL queries over the web of linked data**. In: The Semantic Web-ISWC 2009. Springer Berlin Heidelberg, 2009. p. 293-309.

HILGERT, J. G. A construção do texto "falado" por escrito: a conversação na Internet. Fala e escrita em questão, v. 1, p. 17-55, 2000.

HOLMER, T. **Discourse structure analysis of chat communication**. Language@ Internet, v. 5, n. 9, 2008.

HORIGUCHI, S. et al. **GaChat:A chat system that displays online retrieval information in dialogue text**. Workshop on Visual Interfaces to the Social and the Semantic Web (VISSW2009), IUI2009, Feb 8 2009, Sanibel Island, Florida, USA.

JURAFSKY, Dan et al. **Speech and language processing: An introduction to natural language processing, computational linguistics, and speech recognition**. Upper Saddle River: Prentice Hall, 2000.

KOBILAROV, G. et al. **Media meets semantic web—how the bbc uses dbpedia and linked data to make connections.** In: The semantic web: research and applications. Springer Berlin Heidelberg, 2009. p. 723-737.

LASSILA, O.; HENDLER, J. Embracing. Internet Computing, IEEE, v. 11, n. 3, p. 90-93, 2007.

LASSILA, O.; SWICK, R. R. Resource description framework (RDF) model and syntax specification. W3C Recommendation, 1999

LEHMANN, J. et al. **Dbpedia-a large-scale, multilingual knowledge base extracted from wikipedia.** Semantic Web Journal 2013.

LIKERT, R. **A Technique for the Measurement of Attitudes**. Archives of Psychology, 140, 1–55, 1932.

LOH, Stanley et al. **Recommendation of complementary material during chat discussions**. Knowledge Management & E-Learning: An International Journal (KM&EL), v. 2, n. 4, p. 385-399, 2010.

MANNING, C. D.; RAGHAVAN, P.; SCHÜTZE, H. Introduction to information retrieval. Cambridge: Cambridge University Press, 2008.

MANNING, C. D.; SCHÜTZE, H. Foundations of Statistical Natural Language Processing. MIT Press, 1999, Cambridge, MA, USA.

MIRIZZI, R. et al. **Linked Open Data for content-based recommender systems.** In Proceedings of the 8th International Conference on Semantic Systems (I-SEMANTICS '12), Harald Sack and Tassilo Pellegrini (Eds.). ACM, New York, NY, USA, 1-8.

MORAES, E. L. C. de. **Debatepapo: Sequências conversacionais e visualização do co-texto para compreensão da conversação em bate-papo**. Dissertação de Mestrado, UNIRIO, Setembro 2011

MÜHLPFORDT, M.; WESSNER, M. Explicit referencing in chat supports collaborative learning. In: Proceedings of th 2005 conference on Computer support for collaborative learning: learning 2005: the next 10 years!. International Society of the Learning Sciences, 2005. p. 460-469.

OREILLY, T. What is Web 2.0: Design Patterns and Business Models for the Next Generation of Software. Communications & Strategies, No. 1, p. 17, First Quarter 2007. Disponível em: <a href="http://ssrn.com/abstract=1008839">http://ssrn.com/abstract=1008839</a>

ORENGO, V. M.; HUYCK, C. R. A Stemming Algorithmm for the Portuguese Language. Spire. 2001. p. 186-193.

PASSANT, A. **Dbrec—music recommendations using DBpedia**. The Semantic Web–ISWC 2010. Springer Berlin Heidelberg, 2010. p. 209-224.

PASSANT, A. Measuring Semantic Distance on Linking Data and Using it for Resources Recommendations. AAAI Spring Symposium: Linked Data Meets Artificial Intelligence. 2010.

PASSANT, A.; HEITMANN, B.; HAYES, C. **Using linked data to build recommender systems**. Proceedings of RecSys, New York, USA, 2009.

PIMENTEL, M. G. HiperDiálogo: ferramenta de bate-papo para diminuir a perda de cotexto. Dissertação de Mestrado, NCE-IM-UFRJ, Abril 2002

PIMENTEL, M. G.; FUKS, H.; DE LUCENA, C. J. P. **Co-text loss in textual chat tools**. In: Modeling and Using Context. Springer Berlin Heidelberg, 2003. p. 483-490.

PIMENTEL, M. G.; SAMPAIO, F. F. Comunicografia-uma metodologia para análise de processos de interação que se desenvolvem nas ferramentas de comunicação textual da Internet utilizadas no contexto de Educação a Distância. Revista Brasileira de Informática na Educação, v. 10, n. 1, p. 53-59, 2002.

PORTER, M. F. **An algorithm for suffix stripping**. Program: electronic library and information systems, v. 14, n. 3, p. 130-137, 1980.

RABINER, L. A tutorial on hidden Markov models and selected applications in speech recognition. Proceedings of the IEEE, vol.77, no.2, pp.257,286, Feb 1989

REBEDEA, T.; TRAUSAN-MATU, S.; CHIRU, C. Extraction of Socio-semantic data from chat conversations in collaborative learning communities. In: Times of Convergence. Technologies Across Learning Contexts. Springer Berlin Heidelberg, 2008. p. 366-377.

SACKS, H.; SCHEGLOFF, E. A.; JEFFERSON, G. A simplest systematics for the organization of turn-taking for conversation. Language, p. 696-735, 1974.

SALTON, G.; WONG, A.; YANG, C. A vector space model for automatic indexing. Communications of the ACM, v. 18, n. 11, p. 613-620, 1975.

SHABIR, N.; CLARKE, C. **Using linked data as a basis for a learning resource recommendation system**. 1st International Workshop on Semantic Web Applications for Learning and Teaching Support in Higher Education (SemHE'09), ECTEL'09, Nice, France

SHRIBERG, E. et al. **Identifying Agreement and Disagreement in Conversational Speech: Use of Bayesian Networks to Model Pragmatic Dependencies**. Association for Computational Linguistics Stroudsburg, PA, USA 2004.

SILVA, W. D. C. de M. **Aprimorando o corretor gramatical CoGrOO**. Dissertação (Mestrado em Ciência da Computação) - Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo, 2013.

SMITH, M.; CADIZ, J. J.; BURKHALTER, B. **Conversation trees and threaded chats**. SIGCHI Bulletin, Minneapolis, v. 31, n. 3, p. 21-23, jul, 1999.

STOLCKE, A. et al. Dialogue act modeling for automatic tagging and recognition of conversational speech. Computational linguistics, v. 26, n. 3, p. 339-373, 2000.

TRAUSAN-MATU, S.; REBEDEA, T. **Polyphonic inter-animation of voices in VMT**. In: Studying virtual math teams. Springer US, 2009. p. 451-473.

TUMASJAN, A. et al. **Predicting Elections with Twitter: What 140 Characters Reveal about Political Sentiment**. ICWSM, v. 10, p. 178-185, 2010.

VIERA, G.; FREDDY, A.; VIRGIL, J. **Uma revisão dos algoritmos de radicalização em língua portuguesa**. Information Research, v. 12, n. 3, p. 8, 2006.

APA

### **Apêndices**

# APÊNDICE A – RESPOSTAS DE PERGUNTAS ABERTAS DO QUESTIONÁRIO UTILIZADO NAS DUAS EXPERIMENTAÇÕES

Quais problemas (que geralmente ocorrem em chats com múltiplos usuários) você percebeu que a ferramenta não contribuiu para solucionar?

- Acho que ainda a recomendação está sendo pouco útil, pois reconhece e recomenda ainda termos bastante triviais. Também nao deveria recomendar a mesma palavra mais de uma vez. Talvez alguns termos muito comuns pudessem ser eliminados de alguma forma. Tambem poderia ter um local onde se colocassem termos a recomendar
- O contexto de uma recomendação ficou incoerente
- No meu entender o problema é não só com o reconhecimento de termos, mas de nao termos boas fontes.
- Muitos termos que encontra, nao precisaria recomendar nada. Também acaba sendo muito dependente da riqueza e adequação das fontes para o contexto do Chat.
- Palavras importantes que deveriam ser grifadas não foram.
- Com excessivos usuários, é fácil de se proliferar segmentos de diferentes assuntos.
- Perda de contexto ao longo das conversas. Algumas palavras tem descricao errada,
   levando ao conhecimento erroneo de certas pessoas
- Confusão gerada por muitos usuários digitarem ao mesmo tempo.
- Nem sempre as referências indicada para as palavras foram boas.
- Acho que a ferramenta demora um tempo para fazer as recomendações, provavelmente esse atraso se deve ao processamento que deve ser feito, logo o usuário demora um pouco para se iterar do assunto principal que está sendo debatido. Quando entrei no chat, figuei um pouco perdida sobre o assunto que

- estava sendo debatido e também demorou um certo tempo para aparecer as recomendações, por isso tive que reler todos os debates que já haviam sido feitos.
- O desfoque do assunto não está ligado somente ao tema majoritariamente discutido, mas na intenção do usuário em si. Caso o usuário queira mudar o assunto bruscamente, não será uma recomendação que o fará a pensar o contrário.
- Dificuldade de acompanhar a conversa devido a grande quantidade de usuarios. Os usuários perdidos, pessoas que falam entre si no meio de outras mensagens.
- A ferramenta n\u00e3o contribui para exibir os assuntos discutidos durante o chat de forma diferenciada. Como diversos assuntos foram discutidos, seria interessante a ferramenta exibir esses assuntos de forma diferenciada, para facilitar o entendimento pelos participantes do chat.
- Multiplos assuntos ao mesmo tempo. A conversa nao segue uma linha de raciocinio
- Algumas recomendações não foram feitas e outras eram relevantes para a conversa
- a perda de foco em assuntos, sem esgotá-los
- Acho que ela demora para enviar a mensagem e acaba que quando a mesma chega, fica fora do contexto. Fica difícil direcionar a resposta de algo específico de alguém para algum outro alguém.

### Quais problemas (que geralmente ocorrem em chats com múltiplos usuários) você percebeu que a ferramenta contribuiu para solucionar?

- as recomendações facilitaram muito o entendimento, mas elas vinham muito depois da conversa começar e isso foi um problema
- Algumas palavras foram recomendadas e que contribuíram para o entendimento do assunto.
- As recomendações levarão a um apoio fundamental a quem chegava de fora e não sabia a natureza do assunto tratado em alguns segundos.
- Conceitos que nao sao conhecidos pelos usuarios puderam ser explanados durante a conversacao sem a necessidade de efetuar uma busca, contribuindo para que o usuario nao ficasse perdido em relacao ao que estava sendo dito utilização de termos que as pessoas não se conhecem.
- Nao notei as correlações feitas pela ferramenta, na ficou evidente.

- Quando a pessoa não sabe o assunto, alguns links são mandados corretamente para que a pessoa saiba do que estamos falando.
- Informações sobre os termos que estão sendo discutidos que muitas vezes não é bem conhecido pela pessoa.
- Ajudou a esclarecer o que sao algumas palavras
- Algumas recomendações foram bem uteis para entendimento de alguma referencia feita por algum participante do chat
- A ferramenta ajuda a esclarecer alguns conceitos comentados durante o chat, contribuindo para um melhor entendimento do assunto discutido.
- Os termos são exibidos e recomendados pela própria ferramenta, o que diminui problemas de entendimento e ambiguidades, garantindo que todas as pessoas estão conversando sobre o mesmo termo. Quanto às questões relativas a dúvidas e usuários perdidos, não foi possível observar muito.
- Indicação de termos que foram citados e que em parte um usuário não tinha conhecimento.
- Acho que as principais contribuições da ferramenta é a ajuda ao usuário a se iterar do assunto principal que está sendo debatido no chat, fazendo com que ele fique menos perdido, especialmente quando ele acaba de entrar no chat. Outra contribuição importante é a desambiguação das palavras, a explicação que é feita a partir das palavras recomendadas ajuda bastante o usuário a se iterar melhor sobre o assunto e ter a compreensão certa sobre o que está sendo discutido, melhorando a qualidade do debate no chat.
- Certamente situa muito bem o usuário com as palavras que foram recomendadas.
   Ajudou em dar um contexto sobre o que estava sendo conversado quando eu desconhecia totalmente o assunto e os termos usados.

#### Gostaria de fazer mais algum comentário? Sugestões?

- Acentos não deveriam ser considerados.
- coloquei acima algumas sugestòes
- Um termo muito citado, deve ser recomendado? ou apenas algo que realmente exista a necessidade de desambiguação ou necessidade de obtenção de

- conhecimento sobre o termo? um fórum, separado por assunto, me parece ser uma ferramenta mais adequada
- Ocorreu um bug ao tentar envir e receber uma mensagem de um unico usuário.
   Cliquei em um usuário da lista lateral de usuários, enviei uma mensagem e ela não chegou para o destinatário. O inverso também aconteceu (a mensagem da outra pessoa não chegou pra mim).
- Pude perceber uma certa demora pra que a mensagem enviada por mim aparecesse para os outros usuários do chat.
- Sim, notei haver termos repetidos nas recomendações. Sugiro que se elabore um mecanismo que identifique variantes de cada termo para que se mantenham único independente da grafia utilizada, baseando-se no contexto.
- As sugestões acabaram por servir mais como um histórico resumido dos assuntos abordados do que como sugestões que ampliaram a discussão. Funcionou bem para "tagear" a conversa.
- Deveria se aprimorar o algoritmo de reconhecimento de palavras. Uma modificação que talvez seja simples é reconhecer palavras sem acento e escritas todas com letras minúsculas. Palavras com ortografia errada poderiam ser reconhecidas também. É impossível esperar que num chat de múltiplos usuários as pessoas vão escrever seguindo os padrões ortográficos.
- Algumas das recomendacoes da ferramente foram equivocadas, e por vezes tambem alterou o curso da discussao por escrever na tela onde as mensagens aparecem. A abordagem usando baloes flutuantes me pareceu mais elegante em relacao a esse problema.
- Os 2 problemas que observei foram: Quando clico em cima da recomendação é disparado um alerta com um erro e o usuário é deslogado do site. Acho que esse problema de deslogar o usuário do site pode prejudicar o andamento do debate e muitos usuários podem acabar desistindo de participar do chat, porque foram deslogados. Algumas recomendações aparecem multiplicadas. No final do debate, observei que a recomendação "MPL" estava exposta 4 vezes. Entretanto, a explicação da recomendação esteve igual em todas as repetições.
- A ferramenta deu muitas recomendações repetidas.

- As definições deveriam aparecer mais rapidamente pois, com o tempo grande, vo acaba ficando perdido no assunto ou o assunto muda e ainda não apareceu a definição.
- algumas definições poderiam ser melhores mesmo que o texto ficasse um pouco maior. Identificar quais usuários citaram a palavra (empressão) recomendada.
- Sim, acho que é preciso de alguma forma eliminar recomendações para termos pouco relevantes. Acho díficil, mas pode introduzir um ruido no chat a recomendação de definições de termos muito óbvios.
- Seria legal que fosse possivel responder diretamente uma pessoa.

# APÊNDICE B — DIÁLOGO E RECOMENDAÇÕES GERADAS A PARTIR DA SEGUNDA EXPERIMENTAÇÃO

1 usuario1: oi

2 usuario2: Boa noite

3 usuario3: Olá

4 usuario4: Boa noite

5 usuario3: Hoje vamos conversar sobre a reforma política e o que vocês gostariam que fosse mudado na constituição | recomendação: reforma política - No Brasil, Reforma Política é o nome dado ao conjunto de propostas de emendas constitucionais e PC's da lei eleitoral com fins de tentar melhorar o sistema eleitoral nacional, proporcionando, segundo seus propositores, maior correspondência entre a vontade do eleitor ao votar e o resultado final das urnas

6 usuario4: o voto distrital | recomendação: voto distrital - O voto distrital é, na mídia e nos meios políticos brasileiros, sinônimo de sistema eleitoral de maioria simples. Esse é um sistema em que cada membro do parlamento é eleito individualmente nos limites geográficos de um distrito pela maioria dos votos (simples ou absoluta). Para tanto, o país é dividido em determinado número de distritos eleitorais, normalmente com população semelhante entre si, cada qual elegendo um dos políticos que comporão o parlamento.

7 usuario3: como o usuario4 deu a ideia podemos começar pela proposta de voto distrital, o que vocês acham? | recomendação: voto distrital - O voto distrital é, na mídia e nos meios políticos brasileiros, sinônimo de sistema eleitoral de maioria simples. Esse é um sistema em que cada membro do parlamento é eleito individualmente nos limites geográficos de um distrito pela maioria dos votos (simples ou absoluta). Para tanto, o país é dividido em determinado número de distritos eleitorais, normalmente com população semelhante entre si, cada qual elegendo um dos políticos que comporão o parlamento.

8 usuario2: legal, eu não conhecia a proposta e a recomendação me ajudou

9 usuario5: penso que os distritos com maior poder econômico iriam simplesmente comprar o apoio dos outros...

10 usuario6: como funcionaria esse tipo de voto? Seria dividido por estados?

11 usuario7: acho que a divisão seria por bairros

12 usuario3: eu acho que pela descrição do conceito os distritos podem ser divididos da maneira que os governantes acharem melhor

13 usuario1: tenho dúvidas se essa divisão alteraria alguma coisa

14 usuario8: usuario1, com a divisão por distritos talvez a população conhecesse melhor os parlamentares

15 usuario9: ainda não entendo muito bem como funcionaria esse voto

16 usuario5: o problema é que é difícil dar seu voto para uma pessoa que não se conhece, essa divisão favoreceria a resolução disso

17 usuario3: E o que vocês acham do voto ser obrigatório? / recomendação: voto obrigatório - Voto compulsório é uma prática que requer que os pessoas votem em eleições ou se apresentem em seções eleitorais para justificar sua ausência de um processo eleitoral. Se um eleitor qualificado não se

apresentar a uma seção eleitoral, dependendo do país, ele pode ser punido com multas, serviço comunitário e até mesmo prisão.

18 usuario10: oi Biaa!

19 usuario3: Boa noite usuario10

20 usuario10: o voto obrigatório só vigorará com sucesso quando uma sociedade for suficientemente madura socialmente, ou seja, quando a educação e saúde, por exemplo, estiverem difundidos pelo povo | recomendação: voto obrigatório - Voto compulsório é uma prática que requer que os pessoas votem em eleições ou se apresentem em seções eleitorais para justificar sua ausência de um processo eleitoral. Se um eleitor qualificado não se apresentar a uma seção eleitoral, dependendo do país, ele pode ser punido com multas, serviço comunitário e até mesmo prisão.

21 usuario8: dificultaria a democracia | recomendação: democracia - Democracia é uma forma de governo em que todos os cidadãos elegíveis participam igualmente — diretamente ou através de representantes eleitos — na proposta, no desenvolvimento e na criação de leis. Ela abrange as condições sociais, econômicas e culturais que permitem o exercício livre e igual da autodeterminação política.

22 usuario11: Eu acho que se o voto não for obrigatório a venda de votos será favorecida

23 usuario5: concordo usuario10

24 usuario7: o voto é a arma mais poderosa que temos para mudar alguma coisa nesse país.

25 usuario12: infelizmente, vota-se nas pessoas e os ideias são passados ao largo.

26 usuario12: a educação política vem na prática tb. podem me matar,mas sou a favor ainda do voto obrigatório. | recomendação: voto obrigatório - Voto compulsório é uma prática que requer que os pessoas votem em eleições ou se apresentem em seções eleitorais para justificar sua ausência de um processo eleitoral. Se um eleitor qualificado não se apresentar a uma seção eleitoral, dependendo do país, ele pode ser punido com multas, serviço comunitário e até mesmo prisão.

27 usuario5: usuario12 você vota nos objetivos dos partidos, mas penso que a maioria das pessoa vota nos candidatos sem se importar com o partido

28 usuario13: o pior que estas pessoas são eleitas (efeito do voto obrigatório) simplesmente em função de sua popularidade com a massa | recomendação: voto obrigatório - Voto compulsório é uma prática que requer que os pessoas votem em eleições ou se apresentem em seções eleitorais para justificar sua ausência de um processo eleitoral. Se um eleitor qualificado não se apresentar a uma seção eleitoral, dependendo do país, ele pode ser punido com multas, serviço comunitário e até mesmo prisão.

29 usuario12: desculpe usuario13, acho que isto independe se o voto é obrigatório ou não.

30 usuario9: voce acaba votando em uma bandeira

31 usuario13: usuario12, o voto é uma "arma", para o bem ou para o mau. se a pessoa não sabe usálo acaba favorecendo este tipo de gente. Você sabe quando o voto era em papel, um dos candidatos mais votados era o "macaco tião" (mascote do jardim zoológico do rio). praticamente ele era eleito o governador do rio toda eleição? naquela época as pessoas faziam isso como uma forma de protesto. hoje elas votam no tiririca. o pior que isso gera consequências. | recomendação: tiririca - Francisco Everardo Oliveira Silva (Itapipoca, 1 de maio de 1965), conhecido pelo nome artístico de Tiririca, é um cantor, compositor, humorista, e político brasileiro. Filiado ao Partido da República, Tiririca foi eleito deputado federal por São Paulo, tendo sido o segundo deputado mais votado em toda a história do Brasil.

32 usuario5: você nunca vota numa bandeira

33 usuario11: usuario5, você vota numa bandeira quando vota em branco ou quando digita qualquer coisa e ainda favorece para alianças de 'gangues'

34 usuario6: usuario5, eu voto numa bandeira sim. feita de pessoas, mas numa bandeira e no seu ideal e acho que esse é o caminho

35 usuario13: esse problema de votar pela aparência do candidato não acontece só no brasil, sou a favor do voto facultativo

36 usuario11: também usuario13... isso faria com que somente pessoas interessadas fossem votar

37 usuario5: o voto nos partidos poderia solucionar isso?

38 usuario2: quando a compra de votos. obrigatório ou não, quem vende voto ganha algo com isso então continuariam vendendo.

39 usuario5: mas quem não quer entender é só não votar!

40 usuario5: usuario11 qual o problema de só as pessoas interessadas votarem? Se o resto das pessoas não gostar das medidas passará a se interessar e a votar também

41 usuario7: a maioria vota em candidatos e não em partidos. a lei da ficha limpa deveria ser colocada em pratica. | recomendação: ficha limpa - Ficha Limpa ou Lei Complementar nº. 135 de 2010 é uma legislação brasileira que foi emendada à Lei das Condições de Inelegibilidade ou Lei Complementar nº. 64 de 1990 originada de um projeto de lei de iniciativa popular idealizado pelo juiz Márlon Reis que reuniu cerca de 1,3 milhão de assinaturas com o objetivo de aumentar a idoneidade dos candidatos. A lei torna inelegível por oito anos um candidato que tiver o mandato cassado, renunciar para evitar a cassação ou for condenado por decisão de órgão colegiado (com mais de um juiz), mesmo que ainda exista a possibilidade de recursos.

42 usuario14: voces acham que com voto não obrigatório não haveria "compra" de votos? eu discordo

43 usuario13: testemunhei um caso de sucesso do "ficha limpa". no meu bairro tem um indivíduo que conheço há bastante tempo que se candidatou a deputado estadual nas última eleições. ele é um morador da comunidade do chapéu mangueira, no bairro do leme. pessoalmente eu sei que ele não é uma pessoa de boa índole mas por algumas manobras chegou a conseguir se tornar líder comunitário nesta comunidade. Pois então...aproveitando esta popularidade na comunidade, se lançou político nas últimas eleições. não tenho certeza se ele teve votos suficientes para se eleger mas pelo que soube do desfecho é que ele foi pego no "ficha limpa" e teve sua candidatura impugnada. | recomendação: ficha limpa - Ficha Limpa ou Lei Complementar nº. 135 de 2010 é uma legislação brasileira que foi emendada à Lei das Condições de Inelegibilidade ou Lei Complementar nº. 64 de 1990 originada de um projeto de lei de iniciativa popular idealizado pelo juiz Márlon Reis que reuniu cerca de 1,3 milhão de assinaturas com o objetivo de aumentar a idoneidade dos candidatos. A lei torna inelegível por oito anos um candidato que tiver o mandato cassado, renunciar para evitar a cassação ou for condenado por decisão de órgão colegiado (com mais de um juiz), mesmo que ainda exista a possibilidade de recursos.

44 usuario6: tambpem discordo usuario14. a questão do voto ser ou não orbigatório não é essencial. acho até que o sistema atual tem suas vantagens

45 usuario5: como vou saber se o candidato é honesto

46 usuario13: só varrendo o histórico do candidato. isso se o histórico estiver disponível...

47 usuario5: mas o fato de ele ter a ficha limpa não quer dizer que ele é honesto ou confiável | recomendação: ficha limpa - Ficha Limpa ou Lei Complementar nº. 135 de 2010 é uma legislação brasileira que foi emendada à Lei das Condições de Inelegibilidade ou Lei Complementar nº. 64 de 1990 originada de um projeto de lei de iniciativa popular idealizado pelo juiz Márlon Reis que reuniu

cerca de 1,3 milhão de assinaturas com o objetivo de aumentar a idoneidade dos candidatos. A lei torna inelegível por oito anos um candidato que tiver o mandato cassado, renunciar para evitar a cassação ou for condenado por decisão de órgão colegiado (com mais de um juiz), mesmo que ainda exista a possibilidade de recursos.

48 usuario7: ficha limpa serve pra isso. | recomendação: ficha limpa - Ficha Limpa ou Lei Complementar nº. 135 de 2010 é uma legislação brasileira que foi emendada à Lei das Condições de Inelegibilidade ou Lei Complementar nº. 64 de 1990 originada de um projeto de lei de iniciativa popular idealizado pelo juiz Márlon Reis que reuniu cerca de 1,3 milhão de assinaturas com o objetivo de aumentar a idoneidade dos candidatos. A lei torna inelegível por oito anos um candidato que tiver o mandato cassado, renunciar para evitar a cassação ou for condenado por decisão de órgão colegiado (com mais de um juiz), mesmo que ainda exista a possibilidade de recursos.

49 usuario5: só quer dizer que ele não é bandido

50 usuario12: digo mais usuario5, tem político honesto que é difamado e processado injustamente por interesses políticos. é difícil saber mesmo a verdade.

52 usuario16: oi gente, boa noite!

53 usuario7: se ele for honesto ele é confiável sim.

54 usuario3: boa noite, usuario16

55 usuario5: eu acho muito complicado definir se ele é mesmo confiavel

56 usuario12: acho que só os políticos tem a ganhar agora com o voto facultativo.

57 usuario3: acho que talvez com o plebiscito para a reforma poderíamos votar nessas propostas de emenda a constituição | recomendação: proposta de emenda constitucional - No Brasil, as alterações do texto constitucional têm início com uma propostas de emenda à Constituição (PEC), conforme previsão do artigo 60 da Constituição Federal de 1988. As PECs devem observar uma tramitação especial e pode ser apresentada segundo dicção do artigo 60 da Constituição Federal: I - de um terço, no mínimo, dos membros da Câmara dos Deputados ou do Senado Federal; II - do Presidente da República; III - de mais da metade das Assembléias Legislativas das unidades da Federação, manifestando-se, cada uma delas, pela maioria relativa de seus membros. | plebiscito - O plebiscito (do Lat. plebiscitu - decreto da plebe) era considerado, na Roma antiga, voto ou decreto passados em comício, originariamente obrigatórios apenas para os plebeus. Hoje em dia, o plebiscito é convocado antes da criação da norma (ato legislativo ou administrativo), e são os cidadãos, por meio do voto, que vão aprovar ou não a questão que lhes for submetida.

58 usuario6: mas ainda sim sim usuario3, uma reforma produnda temq que ter corpo, tem que ser pensada no sentido global. por isso acho muito importante a atuação de certos movimentos sociais e partidos de esquerda que colocam um debate ideologico | recomendação: esquerda - Na Esquerda e Direita política, política de esquerda descreve uma visão ou posição específica que aceita ou suporta igualdade social, muitas vezes em oposição a hierarquia social e a desigualdade social. Normalmente envolvendo uma preocupação com os cidadãos que são considerados em desvantagem em relação aos outros e uma suposição de que há desigualdades injustificadas que devem ser reduzidas ou abolidas.

59 usuario14: o relatório resultante do debate no e-democracia, tem pontos interessantes, justificando ou dando a base de cada questão apresentada: voto distrital, coligações, e outros. | recomendação: voto distrital - O voto distrital é, na mídia e nos meios políticos brasileiros, sinônimo de sistema eleitoral de maioria simples. Esse é um sistema em que cada membro do parlamento é eleito individualmente nos limites geográficos de um distrito pela maioria dos votos (simples ou absoluta). Para tanto, o país é dividido em determinado número de distritos eleitorais, normalmente

com população semelhante entre si, cada qual elegendo um dos políticos que comporão o parlamento.

60 usuario5: acredito que um plebiscito com voto não obrigatório nos temas principais seria melhor do que deixar a reforma nas mãos dos interesses dos políticos | recomendação: plebiscito - O plebiscito (do Lat. plebiscitu - decreto da plebe) era considerado, na Roma antiga, voto ou decreto passados em comício, originariamente obrigatórios apenas para os plebeus. Hoje em dia, o plebiscito é convocado antes da criação da norma (ato legislativo ou administrativo), e são os cidadãos, por meio do voto, que vão aprovar ou não a questão que lhes for submetida.

61 usuario12: numa postagem no e-democracia, geraldo azevedo siqueira colocou que: a constituição de 1891 marcou o fim dos privilégios políticos, a constituição inaugural com a reforma de 1934, verão que ela supera em muito nossa atual constituição, principalmente em matéria política. as constituições posteriores, desfiguraram a república original e restauraram os privilégios. ele sugere que o brasil precisaria desconstruir 76 anos de atraso político, e reencontrar sua identidade, para então seguir seu destino. nos últimos 76 anos vimos a aparelhamento do estado no interesse dos partidos políticos, dos próprios políticos e de toda a gente que vive de impostos. isto se comprova pela quantidade de privilégios que foram reinstituídos nas constituições a partir de 1937. | recomendação: constituição - A atual Constituição da República Federativa do Brasil foi promulgada em 5 de outubro de 1988. Ela estabelece o Brasil como um Estado democrático de Direito de estrutura federativa. Em 1993, conforme determinação do texto constitucional, foi realizado um plebiscito para que o povo determinasse a forma de governo, entre monarquia e república, e o sistema de governo, podendo optar entre o presidencialismo e o parlamentarismo. Foi confirmado o regime republicano e o presidencialismo já existentes, junto com a tripartição dos poderes. A República Federativa do Brasil é composta por 26 Estados federados e um Distrito Federal.

62 usuario13: o problema no plebiscito é que o voto é obrigatório, logo pode haver o mesmo problema das eleições, como compra de voto. também, sou a favor do voto facultativo | recomendação: plebiscito - O plebiscito (do Lat. plebiscitu - decreto da plebe) era considerado, na Roma antiga, voto ou decreto passados em comício, originariamente obrigatórios apenas para os plebeus. Hoje em dia, o plebiscito é convocado antes da criação da norma (ato legislativo ou administrativo), e são os cidadãos, por meio do voto, que vão aprovar ou não a questão que lhes for submetida.

63 usuario5: a única ferramenta que temos para regular o comportamento dos políticos hoje são as leis, concordo que com o plebiscito poderíamos ter participação e quiça apertar mais o cinto dos políticos... | recomendação: plebiscito - O plebiscito (do Lat. plebiscitu - decreto da plebe) era considerado, na Roma antiga, voto ou decreto passados em comício, originariamente obrigatórios apenas para os plebeus. Hoje em dia, o plebiscito é convocado antes da criação da norma (ato legislativo ou administrativo), e são os cidadãos, por meio do voto, que vão aprovar ou não a questão que lhes for submetida.

64 usuario13: acho que essa é a ideia do plebiscito ver a real opinião das pessoas, não opiniões compradas | recomendação: plebiscito - O plebiscito (do Lat. plebiscitu - decreto da plebe) era considerado, na Roma antiga, voto ou decreto passados em comício, originariamente obrigatórios apenas para os plebeus. Hoje em dia, o plebiscito é convocado antes da criação da norma (ato legislativo ou administrativo), e são os cidadãos, por meio do voto, que vão aprovar ou não a questão que lhes for submetida.

65 usuario6: usuario5, pelbiscito é muito bom, mas não pra todos os casos. Em especial, quando se trata de direito de minorias. por exemplo, hoje, um pelbiscito proibiria o casamento homossexual

66 usuario15: plebiscito é para coisa pontual...não espere que a população voto consciente sem entender do que se trata o plebiscito...não estou falando só de classe baixa, ok? | recomendação: plebiscito - O plebiscito (do Lat. plebiscitu - decreto da plebe) era considerado, na Roma antiga, voto

ou decreto passados em comício, originariamente obrigatórios apenas para os plebeus. Hoje em dia, o plebiscito é convocado antes da criação da norma (ato legislativo ou administrativo), e são os cidadãos, por meio do voto, que vão aprovar ou não a questão que lhes for submetida.

67 usuario6: não dá pra acreditar nele enquanto não houver uma reforma das midias também

68 usuario6: o poder delas é muito grande e não há qualquer controle social sobre elas

69 usuario5: usuario6 como você acha que um sistema por voto majoritária irá satisfazer as minorias? | recomendação: voto majoritário - Em um sistema eleitoral majoritário (português brasileiro) ou maioritário (português europeu), o vitorioso surge por maioria relativa (ou pluralidade), que representa a maior minoria, ou por maioria absoluta. Foi o primeiro princípio representativo surgido na história ocidental, a partir das práticas da Igreja Católica ainda na Idade Média, para definir representantes de suas várias circunscrições administrativas espalhadas pela Europa. Posteriormente, tal princípio foi adotado na Inglaterra para a escolha de representantes políticos das regiões do país. Atualmente, está em vigor em países como Brasil, Portugal, EUA, Canadá, Inglaterra, França, Austrália e Nova Zelândia

70 usuario16: controlar a mídia é polêmico... ela exerce o qurto poder e está acima dos outros três, como fazer?

71 usuario14: uma mudança de cultura pode acontecer. acho que temos muitos mecanismos de envolvimento hoje

72 usuario16: a imaturidade ainda é muito grande nesse país.

73 usuario5: usuario6 você disse que um plebiscito é prejudicial para as minorias, eleição por voto majoritária é bom para as minorias? | recomendação: plebiscito - O plebiscito (do Lat. plebiscitu - decreto da plebe) era considerado, na Roma antiga, voto ou decreto passados em comício, originariamente obrigatórios apenas para os plebeus. Hoje em dia, o plebiscito é convocado antes da criação da norma (ato legislativo ou administrativo), e são os cidadãos, por meio do voto, que vão aprovar ou não a questão que lhes for submetida. | voto majoritário - Em um sistema eleitoral majoritário (português brasileiro) ou maioritário (português europeu), o vitorioso surge por maioria relativa (ou pluralidade), que representa a maior minoria, ou por maioria absoluta. Foi o primeiro princípio representativo surgido na história ocidental, a partir das práticas da Igreja Católica ainda na Idade Média, para definir representantes de suas várias circunscrições administrativas espalhadas pela Europa. Posteriormente, tal princípio foi adotado na Inglaterra para a escolha de representantes políticos das regiões do país. Atualmente, está em vigor em países como Brasil, Portugal, EUA, Canadá, Inglaterra, França, Austrália e Nova Zelândia

74 usuario6: usuario16, imaturiidade é uma coisas que não existe em politica. existe falata de organização e de consciencia, mas imaturidade não

75 usuario6: só disse que um plebicito pode ser muito ruim para questões de minoria

76 usuario16: oi tiago, sem problema.

77 usuario14: acho que nosso desconhecimento em política, sobre a proposta dos partidos e legislação é um sintoma de que precisamos mudar na raiz. as discussões se dão em ponto pequeno e sem continuidade, envolvendo apenas os diretamente interessados nao temos a cultura da participação.

78 usuario6: claro que a ignorancia atrapalha, mas nem todo iletrado é um analfabeto politico. a classe media é que vive achando que saber o que é bom e o que não é

79 usuario7: o país só muda com a participação direta da sociedade.

80 usuario2: o problema das leis é a cultura de não cumprí-las. o famoso "jeitinho brasileiro" que está aí pra tudo.

81 usuario10: que diga seu cesar maia e garotinho | recomendação: cesar maia - Cesar Epitácio Maia é um economista e político brasileiro, ex-prefeito do Rio de Janeiro pelo partido político Democratas. É o prefeito da cidade do Rio de Janeiro que permaneceu mais tempo no cargo (12 anos). É pai de Daniela Maia e Rodrigo Maia, este deputado federal e presidente nacional do Democratas, além de primo do político José Agripino Maia. Cesar Maia foi cotado para ser candidato a presidente da República pelo Democratas (DEM) para o pleitos de 2006 e de 2010. Em 2006, renunciou a précandidatura. Já em 2010, foi escolhido pela coligação PV, PSDB, PPS e DEM como candidato ao Senado Federal. Acabou sendo derrotado no referido pleito, alcançando o quarto lugar, desempenho considerado muito abaixo do esperado, se comparado com seu histórico político no Rio de Janeiro. | garotinho - Anthony William Matheus de Oliveira, conhecido como Anthony Garotinho, é um radialista e político brasileiro. Foi o 58º governador do Rio de Janeiro e candidato à presidência da república em 2002. Garotinho tornou-se conhecido nacionalmente ao longo dos dois mandatos em que foi o prefeito de Campos, sua cidade natal(na época foi considerado o melhor prefeito do Brasil).

82 usuario14: vejam que voltamos sempre ao ponto de "criar regras" mas não acho que este seja nosso maior problema: precisamos de mecanismso efetivos para que se cumpram as muitas regras que temos

83 usuario5: concordo com a usuario14, o que estamos tentando fazer é criar regras que não tenham nenhuma brecha para os aproveitadores, e já fazemos isso há algum tempo, será que é a solução?

84 usuario6: também acho que a participação no pais precisa aumentar e muito

85 usuario16: usuario6, veja quem foram (e quantas foram) as pessoas que participaram das mobilizações em 2013.

86 usuario6: foram muitas e bem diversificadas usuario16. faltou organização politica, mas decadas de desmobilização tem seus efeitos negativos

87 usuario6: o banlanço entre os poderes é outro problema. acho que no brasl o legislativo é inútil e o executivo muito inchado | recomendação: executivo - O Poder Executivo do Brasil é um dos três poderes existentes no país. É também o conjunto dos órgãos e autoridades públicas aos quais a Constituição Federal brasileira (a atual é de 1988) atribui a função administrativa e adota os princípios da soberania popular e da representação, segundo os quais o poder político, teoricamente, pertence ao povo e é exercido em nome deste por órgãos constitucionalmente definidos (art. 1º, parágrafo único) | legislativo - O Poder Legislativo do Brasil é um dos poderes constituídos pelo topolino chefe do país. Ele é exercido, no âmbito federal, desde 1891, pelo Congresso Nacional, que se compõe da Câmara dos Deputados e do Senado Federal, compostos, respectivamente, por deputados federais e senadores.

88 usuario3: não entendi

89 usuario6: por exemplo, o orçamento, hoje, na prática, não está nas mãos do legislativo | recomendação: legislativo - O Poder Legislativo do Brasil é um dos poderes constituídos pelo topolino chefe do país. Ele é exercido, no âmbito federal, desde 1891, pelo Congresso Nacional, que se compõe da Câmara dos Deputados e do Senado Federal, compostos, respectivamente, por deputados federais e senadores.

90 usuario5: concordo usuario6

91 usuario15: usuario6 é verdade. mais é conveniente não estar nas mãos deles, pois eles teriam muito trabalho...

92 usuario6: o problema, usuario15, é que o legislativo tem um caráter essencialmente mais democratico que o executivo. acho que um legislativo forte, aliado a formas de escolha direta da população, inclusive plebescitos, podem ser muito interessantes | recomendação: executivo - O Poder Executivo do Brasil é um dos três poderes existentes no país. É também o conjunto dos órgãos e

autoridades públicas aos quais a Constituição Federal brasileira (a atual é de 1988) atribui a função administrativa e adota os princípios da soberania popular e da representação, segundo os quais o poder político, teoricamente, pertence ao povo e é exercido em nome deste por órgãos constitucionalmente definidos (art. 1º, parágrafo único) | legislativo - O Poder Legislativo do Brasil é um dos poderes constituídos pelo topolino chefe do país. Ele é exercido, no âmbito federal, desde 1891, pelo Congresso Nacional, que se compõe da Câmara dos Deputados e do Senado Federal, compostos, respectivamente, por deputados federais e senadores.

93 usuario15: o legislativo deveria tomar as rédeas dos projetos de longo prazo e deixar o executivo cuidando só dos problemas do dia-a-dia... | recomendação: executivo - O Poder Executivo do Brasil é um dos três poderes existentes no país. É também o conjunto dos órgãos e autoridades públicas aos quais a Constituição Federal brasileira (a atual é de 1988) atribui a função administrativa e adota os princípios da soberania popular e da representação, segundo os quais o poder político, teoricamente, pertence ao povo e é exercido em nome deste por órgãos constitucionalmente definidos (art. 1º, parágrafo único) | legislativo - O Poder Legislativo do Brasil é um dos poderes constituídos pelo topolino chefe do país. Ele é exercido, no âmbito federal, desde 1891, pelo Congresso Nacional, que se compõe da Câmara dos Deputados e do Senado Federal, compostos, respectivamente, por deputados federais e senadores.

94 usuario6: concordo usuario15. acho que educação, por exemplo, é uma típica área que é melhor gerida pelo legislativo. | recomendação: legislativo - O Poder Legislativo do Brasil é um dos poderes constituídos pelo topolino chefe do país. Ele é exercido, no âmbito federal, desde 1891, pelo Congresso Nacional, que se compõe da Câmara dos Deputados e do Senado Federal, compostos, respectivamente, por deputados federais e senadores.

95 usuario5: como o legislativo iria gerir algo, seu papel não é criar leis?

96 usuario6: infelizmente, no brasil, o legislativo é feito de ruralista e pela bancada evangélica, tão ruim, que não dá nem pra defender uma constituinte, já que o provável, é que os constituintes iriam regredir o carater democratico | recomendação: constituinte - Uma assembleia constituinte (FO 1943: assembléia constituinte) é um organismo colegiado que tem como função redigir ou reformar a constituição, a ordem político-institucional de um Estado, sendo para isso dotado de plenos poderes ou poder constituinte, ao qual devem submeter-se todas as instituições públicas. | bancada evangélica - Na política do Brasil, Frente Parlamentar Evangélica, ou simplesmente bancada evangélica, é um termo aplicado a uma frente parlamentar do Congresso Nacional do Brasil composta por políticos evangélicos de partidos distintos que se articulam de forma contrária a questões como a legalização do aborto, o casamento entre pessoas do mesmo sexo, a criminalização da discriminação contra homossexuais e a prática da eutanásia, por exemplo. | ruralista - A União Democrática Ruralista (UDR) é uma entidade associativa brasileira que reúne proprietários rurais e tem como objetivo declarado "a preservação do direito de propriedade e a manutenção da ordem e respeito às leis do País".

97 usuario15: o legislativo deveria tomar as rédias, ou seja controlar o rumo....

98 usuario15: somos uma nação rural...a maior parte do pib vem do campo... | recomendação: pib - O produto interno bruto (PIB) representa a soma (em valores monetários) de todos os bens e serviços finais produzidos numa determinada região (quer sejam países, estados ou cidades), durante um período determinado (mês, trimestre, ano, etc). O PIB é um dos indicadores mais utilizados na macroeconomia com o objetivo de mensurar a atividade econômica de uma região.

99 usuario12: pessoal, sei que temos que ter liberdade, mas 32 partidos políticos é muito para qq brasileiro. pra entender os objetivos de cada um é difícil.

100 usuario17: tenho minhas dúvidas em considerar somente ruralistas e bancada evangélica. | recomendação: bancada evangélica - Na política do Brasil, Frente Parlamentar Evangélica, ou

simplesmente bancada evangélica, é um termo aplicado a uma frente parlamentar do Congresso Nacional do Brasil composta por políticos evangélicos de partidos distintos que se articulam de forma contrária a questões como a legalização do aborto, o casamento entre pessoas do mesmo sexo, a criminalização da discriminação contra homossexuais e a prática da eutanásia, por exemplo. | ruralista - A União Democrática Ruralista (UDR) é uma entidade associativa brasileira que reúne proprietários rurais e tem como objetivo declarado "a preservação do direito de propriedade e a manutenção da ordem e respeito às leis do País".

101 usuario10: 32 é absurdo... ngm entende nada

102 usuario6: então, não só o mecanismo de escolha do executivo é totalmente tosoco, porque as coligações são de 20 partidos, mas também não sabemos que estamos escolhendo para o legislativo | recomendação: coligações - Uma coligação ou coalizão política é um pacto entre dois ou mais partidos políticos, normalmente de ideias afins, para governar um país, uma região ou outra entidade administrativa. Na maioria dos regimes democráticos, as coligações são permitidas e originam-se quando um só partido ou grupo político não tem apoio suficiente na câmara legislativa correspondente (parlamento, senado ou ambos), como consequência dos resultados eleitorais. | executivo - O Poder Executivo do Brasil é um dos três poderes existentes no país. É também o conjunto dos órgãos e autoridades públicas aos quais a Constituição Federal brasileira (a atual é de 1988) atribui a função administrativa e adota os princípios da soberania popular e da representação, segundo os quais o poder político, teoricamente, pertence ao povo e é exercido em nome deste por órgãos constitucionalmente definidos (art. 1º, parágrafo único) | legislativo - O Poder Legislativo do Brasil é um dos poderes constituídos pelo topolino chefe do país. Ele é exercido, no âmbito federal, desde 1891, pelo Congresso Nacional, que se compõe da Câmara dos Deputados e do Senado Federal, compostos, respectivamente, por deputados federais e senadores.

103 usuario6: o executivo, sempre tende a priorizar as politicas de curto prazo, porque não existe um legsllativo eficiente comprometido com o longo prazo. | recomendação: executivo - O Poder Executivo do Brasil é um dos três poderes existentes no país. É também o conjunto dos órgãos e autoridades públicas aos quais a Constituição Federal brasileira (a atual é de 1988) atribui a função administrativa e adota os princípios da soberania popular e da representação, segundo os quais o poder político, teoricamente, pertence ao povo e é exercido em nome deste por órgãos constitucionalmente definidos (art. 1º, parágrafo único) | legislativo - O Poder Legislativo do Brasil é um dos poderes constituídos pelo topolino chefe do país. Ele é exercido, no âmbito federal, desde 1891, pelo Congresso Nacional, que se compõe da Câmara dos Deputados e do Senado Federal, compostos, respectivamente, por deputados federais e senadores.

104 usuario5: com a reforma poderíamos eliminar leis antigas que não fazem mais sentido. Principalmente as da época da ditadura | recomendação: ditadura - Regime ou Ditadura militar no Brasil foi o regime autoritário que governou o país de 1º de abril de 1964 até 15 de março de 1985. A implantação da ditadura começou com o Golpe de 1964, quando as Forças Armadas do Brasil derrubaram o governo do presidente constitucional João Goulart e terminou quando José Sarney assumiu o cargo de presidente.

105 usuario6: verdade usuario5, podemos eliminar o que não queremos

106 usuario5: todas os dados das contas do governo deveriam estar abertos, alguém sabe a situação hoje?

107 usuario3: acho que eles estão caminhando com o portal da transparência | recomendação: portal da transparência - Regula o acesso a informações previsto no inciso XXXIII do art. 5º, no inciso II do § 3º do art. 37 e no § 2º do art. 216 da Constituição Federal; altera a Lei nº 8.112, de 11 de dezembro de 1990; revoga a Lei nº 11.111, de 5 de maio de 2005, e dispositivos da Lei nº 8.159, de 8 de janeiro de 1991; e dá outras providências.

- 108 usuario14: o proprio pessoal de ti do congresso nacional tem feito um bom trabalho | recomendação: congresso nacional O Congresso Nacional é o órgão constitucional que exerce, no âmbito federal, as funções do Poder Legislativo, quais sejam, elaborar/aprovar leis e fiscalizar o Estado Brasileiro (suas duas funções típicas), bem como administrar e julgar (funções atípicas).
- 109 usuario3: pessoal, só lembrando a quem entrou agora, passando o mouse em cima das palavras do lado esquerdo vocês conseguem ver as recomendações
- 110 usuario14: a abertura de dados vai contribuir para coibir muitos excessos. eu acredito que temos muito força para mudar, apenas nao nos envolvemos os suficiente, em todas as instancias onde temos representatividade
- 111 usuario10: outra coisa é a banalização dos políticos pulando de partido em partido!
- 112 usuario5: mas os partidos obedecem uma lei rígida para criação, vide o da marina silva que não conseguiu sair do papel | recomendação: marina silva Maria Osmarina Marina Silva Vaz de Lima (Rio Branco, 8 de fevereiro de 1958) é uma ambientalista, historiadora, pedagoga e política brasileira recentemente filiada ao Partido Socialista Brasileiro.
- 113 usuario11: o maior problema é o pouco caso que os políticos fazem com qualquer coisa que dure mais de 4 anos ou quando estão no final do seu mandato | recomendação: mandato Mandato é uma palavra usada para definir o espaço de tempo público ou privado que um ocupante, nomeado ou não, concursado ou não e designado que ocupe em caráter provisório um cargo ou função pública.
- 114 usuario12: como eu tinha falado, o problema é a quantidade de partidos que existe hoje também
- 115 usuario3: usuario12, acho que por isso foi votada uma pec para desfiliação partidária | recomendação: desfiliação partidária PL 5652/2013: Dispõe sobre a perda de cargo eletivo em decorrência de desfiliação partidária, nos termos que específica.
- 116 usuario14: a questão é essa mesma usuario3, garantir um minimo de coerência com o que foi proposto quando da criação do partido
- 117 usuario8: criar cláusulas mais rígidas não só para criação, mas para a

#### manutenção do partido é um começo

- 118 usuario3: com relação a isso também existe uma PEC que propoe que o financiamento de campanhas seja exclusivamente público | recomendação: financiamento de campanhas seja exclusivamente público PL 5177/2013: Dispõe sobre o sistema eleitoral para as eleições proporcionais, alterando a Lei nº 4.737, de 15 de julho de 1965 (Código Eleitoral), a Lei nº 9.096, de 19 de setembro de 1995 (Lei dos Partidos Políticos), e a Lei nº 9.504, de 30 de setembro de 1997 (Lei das Eleições) e sobre o financiamento exclusivamente público de campanha eleitoral, alterando o art. 38 da Lei nº 9.096, de 19 de setembro de 1995.
- 119 usuario5: sou a favor da lei da ficha limpa. vocês acham que o financiamaneto exlusivamente público das campanhas tiraria a influência econômica das eleições. hoje ganha eleição quem tem mais dinheiro... | recomendação: ficha limpa Ficha Limpa ou Lei Complementar nº. 135 de 2010 é uma legislação brasileira que foi emendada à Lei das Condições de Inelegibilidade ou Lei Complementar nº. 64 de 1990 originada de um projeto de lei de iniciativa popular idealizado pelo juiz Márlon Reis que reuniu cerca de 1,3 milhão de assinaturas com o objetivo de aumentar a idoneidade dos candidatos
- 120 usuario5: vi um estudo americano que adivinhava os vencedores das disputas eleitorais analisando o rosto dos candidatos, o marketing que vence as campanhas!
- 121 usuario6: os partidecos de aluguél gazem aliança o tempo todo pra aumentar o tempo de tv de partidos grandes e bargnhar com isso

122 usuario5: E será que com a questão do financiamento acabariamos com os partidos de aluguel? | recomendação: partidos de aluguel? - Um partido de aluguel é um partido político usado por um partido mais forte com finalidade estratégica e que, em alguns casos, não visa à vitória nas urnas. Geralmente são partidos nanicos, com muito pouca expressão eleitoral, os que se submetem a tal finalidade. Na maioria das vezes, seus dirigentes (municipais, regionais e nacionais) recebem benefícios pessoais por compactuar com a submissão dessas entidades a outras.

123 usuario6: não, mas acho que proibir coligações já é um duro golpe nos partidos de aluguél usuario5 | recomendação: coligações - Uma coligação ou coalizão política é um pacto entre dois ou mais partidos políticos, normalmente de ideias afins, para governar um país, uma região ou outra entidade administrativa. | partidos de aluguél? - Um partido de aluguel é um partido político usado por um partido mais forte com finalidade estratégica e que, em alguns casos, não visa à vitória nas urnas. Geralmente são partidos nanicos, com muito pouca expressão eleitoral, os que se submetem a tal finalidade. Na maioria das vezes, seus dirigentes (municipais, regionais e nacionais) recebem benefícios pessoais por compactuar com a submissão dessas entidades a outras.

124 usuario16: é, a imprensa apoia quem paga mais...

125 usuario16: será que as propostas de reforma vão realmente incluir todas as essas propostas? Alguém sabe?

126 usuario3: temos algumas propostas, foi criado um grupo de trabalho da câmara dos deputados que está definindo algumas pecs e projetos de lei | recomendação: câmara dos deputados - Câmara dos Deputados é o nome dado às câmaras baixas de congressos ou parlamentos bicamerais. Este nome também é empregado para designar algumas casas legislativas de países com sistema unicameral. | projetos de lei - Um projeto de lei é um tipo de proposta normativa submetida à deliberação de um órgão legislativo, com o objetivo de produzir uma lei. Normalmente, um projeto de lei depende ainda da aprovação ou veto pelo Poder Executivo antes de entrar em vigor.

127 usuario14: é nesta linha o que vem sendo proposto no rela torio da reforma, usuario16

128 usuario6: obrigado usuario14, vou dar uma olhada

129 usuario3: também gostei da proposta do psol | recomendação: psol - O Partido Socialismo e Liberdade (PSOL) é um partido político brasileiro de esquerda fundado em 2004. Obteve registro definitivo na Justiça Eleitoral no dia 15 de setembro de 2005. Seu número eleitoral é 50.

130 usuario5: qual é a proposta do psol | recomendação: psol - O Partido Socialismo e Liberdade (PSOL) é um partido político brasileiro de esquerda fundado em 2004. Obteve registro definitivo na Justiça Eleitoral no dia 15 de setembro de 2005. Seu número eleitoral é 50.

131 usuario13: mas sabemos que isto é praticamente impossível uma vez que uma proposta dessas sempre passará pelo crivo destes indivíduos

132 usuario6: ela é extensa usuario5, mas tem vários pontos que passamos aqui. financiamento exclusivamente público de campanha, proibição de coligações | recomendação: coligações - Uma coligação ou coalizão política é um pacto entre dois ou mais partidos políticos, normalmente de ideias afins, para governar um país, uma região ou outra entidade administrativa. | financiamento exclusivamente público de campanha - PL 5177/2013: Dispõe sobre o sistema eleitoral para as eleições proporcionais, alterando a Lei nº 4.737, de 15 de julho de 1965 (Código Eleitoral), a Lei nº 9.096, de 19 de setembro de 1995 (Lei dos Partidos Políticos), e a Lei nº 9.504, de 30 de setembro de 1997 (Lei das Eleições) e sobre o financiamento exclusivamente público de campanha eleitoral, alterando o art. 38 da Lei nº 9.096, de 19 de setembro de 1995.

133 usuario3: acho que com as manifestações poderíamos conseguir algumas mudanças. O ruim foi não ter propostas muito definidas

134 usuario5: penso que manifestação não exige propostas, apenas é uma "manifestação" de que estamos insatisfeitos quem tem que apresentar alternativas são os políticos, esse é o trabalho deles

135 usuario14: hum, mesmo? nao vi continuidade nem propostas...

136 usuario14: cf 1934

137 usuario14: antonio brito

138 usuario3: em letra maiuscula é mais facil reconhecer, antônio brito | recomendação: antonio brito - António Brito da Silva é um jurista português que desempenhou o cargo de Presidente da CP - Comboios de Portugal, EP entre 1995 e 1997.

139 usuario14: (desculpa, pessoal, estava fazendo uns testes)

140 usuario3: o que voces acham, por exemplo, do programa do bolsa família? | recomendação: bolsa família - O Programa Bolsa Família (PBF) é um programa do Governo Lula (2003) de transferência direta de renda com condicionalidades, que beneficia famílias em situação de pobreza e de extrema pobreza, criado para integrar e unificar ao Fome Zero os programas implantados no Governo FHC: o Bolsa Escola, o Auxílio Gás, o Bolsa Alimentação e o Cartão Alimentação

141 usuario1: acho que o bolsa família foi uma das coisas mais importantes que aconteceram no brasil nos últimos anos. tornou visíveis cerca de 50 milhões de pessoas, tornou-os mais cidadãos. essa talvez seja a maior conquista | recomendação: bolsa família - O Programa Bolsa Família (PBF) é um programa do Governo Lula (2003) de transferência direta de renda com condicionalidades, que beneficia famílias em situação de pobreza e de extrema pobreza, criado para integrar e unificar ao Fome Zero os programas implantados no Governo FHC: o Bolsa Escola, o Auxílio Gás, o Bolsa Alimentação e o Cartão Alimentação

142 usuario5: acredito que deveria se ter uma compensação para o tempo que se dedica no governo sim, não sou a favor de não ganharem nada

143 usuario3: usuario5, o bolsa família acaba fazendo com que as pessoas prefiram não trabalhar | recomendação: bolsa família - O Programa Bolsa Família (PBF) é um programa do Governo Lula (2003) de transferência direta de renda com condicionalidades, que beneficia famílias em situação de pobreza e de extrema pobreza, criado para integrar e unificar ao Fome Zero os programas implantados no Governo FHC: o Bolsa Escola, o Auxílio Gás, o Bolsa Alimentação e o Cartão Alimentação

144 usuario6: o rpograma tem sido desvirtuado por argumentos preconceituosos e não embasados em estatsíticas. o certo, é que ele ajudou a superar a grave seca dos últimos anos

145 usuario8: para min o programa acaba criando o voto de cabresto | recomendação: voto de cabresto - O voto de cabresto é um sistema tradicional de controle de poder político através do abuso de autoridade, compra de votos ou utilização da máquina pública. É um mecanismo muito recorrente nos rincões mais pobres do Brasil como característica do coronelismo.

146 usuario6: usuario8. o bolsa pode até criar voto de cabreto em alguns lugares, mas não é relevante isso - que aliás já iria acontecer de qualquer forma - tanto que a nova classe média não é petista

147 usuario6: o bolsa familía pode tentar dar uma vida mais digana a uma parcela importante da populção que gera muita pressão social e teria pouquissima chance de se inserir no processo produtivo, porque o desemprego é estrutural no sistema | recomendação: bolsa família - O Programa Bolsa Família (PBF) é um programa do Governo Lula (2003) de transferência direta de renda com condicionalidades, que beneficia famílias em situação de pobreza e de extrema pobreza, criado para integrar e unificar ao Fome Zero os programas implantados no Governo FHC: o Bolsa Escola, o Auxílio Gás, o Bolsa Alimentação e o Cartão Alimentação

148 usuario3: mas ele deveria ser um programa temporário

149 usuario4: a questão não é essa

150 usuario2: eu acho que a retirada do benefício acaba fazendo com que as famílias não consigam se manter novamente

151 usuario5: usuario2 quando uma pessoa consegue renda suficiente para manter sua família, retirar r\$ 152 não vai fazê-la ficar pobre novamente. apesar de acreditar que irá demorar gerações

152 usuario 10: mais sem a base da eduação não tem como escalar o conhecimento

153 usuario10: certamente esse cenário seria menos danoso do que é hoje.

154 usuario6: é o tipo de investimento que não dá retorno em 4 anos

155 usuario6: leonoes..que educação? educação de mercado não é a questão. sabemos muito bem que pessoas altamente escolarizadas podem ser avessas a politica. a educação politica não se aprende necessariamente na escola

156 usuario10: é uma coisa acaba puxando a outra

157 usuario3: eu acho que não vai acabar nunca o programa

158 usuario11: acho que devia ter algo a mais em troca ao invés de mandar o filho para uma escola que naturalmente é ruim

159 usuario15: deveria ter investimento em todas as áreas básicas de longo prazo...segurança, projetos de infraestrutura etc..

160 usuario13: absurdo os salários que estes indivíduos recebem! a reforma deveria começar por aí!

161 usuario6: não sei não. tive diversos professores extremamente inteligentes, mas que em termos politicos só faltavam relinchar

162 usuario5: concordo usuario10

163 usuario6: não que a punição tenha que ser abandonada, mas os males - corrupção, etc - não são escoolhas individuais, são um projeto de poder. por exmplo, a alston e ganhou muito com a corrupção de politocs do psdb / recomendação: corrupção - Corrupção o ato ou efeito de se corromper, oferecer algo para obter vantagem em negociata onde favorece uma pessoa e prejudica outra. Tirar vantagem do poder atribuido. | psdb - Partido da Social Democracia Brasileira (PSDB) é um partido político brasileiro. Foi fundado em 25 de junho de 1988 por importantes figuras do cenário político brasileiro, como o ex-presidente (à época senador) Fernando Henrique Cardoso. Seu símbolo é um tucano nas cores azul e amarela e, por esta razão, seus membros são chamados de "tucanos". Em 2012 foi o segundo partido em número de prefeitos eleitos no Brasil, superado apenas pelo PMDB

164 roma: não sabia disso usuario6

165 usuario8: também temos o problema da população que não entende o funcionamento dos parlamentares e não sabem votar

166 roma: o que você quis dizer com isso usuario8?

167 usuario5: acredito que ele se refere a pessoas que não sabem o que um vereador faz serem obrigadas a votar em um

168 usuario5: eu não consigo distinguir as bandeiras do pt psdb e dem por exemplo, acredito que são bandeiras fictícias | recomendação: pt - O Partido dos Trabalhadores (PT) é um partido político brasileiro. Fundado em 1980, é um dos maiores e mais importantes movimentos de esquerda da América do Sul. Com 1 549 180 filiados, o PT é o segundo maior partido político do Brasil, atrás

apenas do Partido do Movimento Democrático Brasileiro (PMDB) | psdb - Partido da Social Democracia Brasileira (PSDB) é um partido político brasileiro. Foi fundado em 25 de junho de 1988 por importantes figuras do cenário político brasileiro, como o ex-presidente (à época senador) Fernando Henrique Cardoso. Seu símbolo é um tucano nas cores azul e amarela e, por esta razão, seus membros são chamados de "tucanos". Em 2012 foi o segundo partido em número de prefeitos eleitos no Brasil, superado apenas pelo PMDB

169 usuario8: igualar as bandeiras pt = psdb é a mesma coisa que dizer que nosso país permaneceu igual de 94-2002 para 2003-atualmente | recomendação: pt - O Partido dos Trabalhadores (PT) é um partido político brasileiro. Fundado em 1980, é um dos maiores e mais importantes movimentos de esquerda da América do Sul. Com 1 549 180 filiados, o PT é o segundo maior partido político do Brasil, atrás apenas do Partido do Movimento Democrático Brasileiro (PMDB) | psdb - Partido da Social Democracia Brasileira (PSDB) é um partido político brasileiro. Foi fundado em 25 de junho de 1988 por importantes figuras do cenário político brasileiro, como o ex-presidente (à época senador) Fernando Henrique Cardoso. Seu símbolo é um tucano nas cores azul e amarela e, por esta razão, seus membros são chamados de "tucanos". Em 2012 foi o segundo partido em número de prefeitos eleitos no Brasil, superado apenas pelo PMDB

170 usuario6: sim usuario8, a maioria, na verdade, os que não são são em geram pequenos partidos de esquerda tipo pco | recomendação: esquerda - Na Esquerda e Direita política, política de esquerda descreve uma visão ou posição específica que aceita ou suporta igualdade social, muitas vezes em oposição a hierarquia social e a desigualdade social. | pco - Partido da Causa Operária (PCO) é um partido político brasileiro de extrema-esquerda.

171 usuario17: Será que não conseguiriamos seguir sem partidos, a população votaria nos ideais dos políticos

172 usuario11: realmente usuario17, não dava para ficar sem partido? Vejo esses partidos, muitas vezes, como gangues

173 usuario13: muitos se candidatam simplesmente para tirar proveito das absurdas regalias do governo para estes indivíduos que não sabem nem porque e para que estão lá. vide tiririca, lacraia, palhaço pipoca,... | recomendação: tiririca - Francisco Everardo Oliveira Silva (Itapipoca, 1 de maio de 1965), conhecido pelo nome artístico de Tiririca, é um cantor, compositor, humorista, e político brasileiro. Filiado ao Partido da República, Tiririca foi eleito deputado federal por São Paulo, tendo sido o segundo deputado mais votado em toda a história do Brasil.

174 usuario13: é mesmo usuario11, uma prova disso é o número de candidatos "artistas" que vêm sendo eleitos

175 usuario5: concordo usuario11, por isso os países fazem leis sem fim desde que a democracia foi inventada | recomendação: democracia - Democracia é uma forma de governo em que todos os cidadãos elegíveis participam igualmente — diretamente ou através de representantes eleitos — na proposta, no desenvolvimento e na criação de leis. Ela abrange as condições sociais, econômicas e culturais que permitem o exercício livre e igual da autodeterminação política.

176 usuario6: nolssco, bom ponto. talvez o foco tenha que ser mudado então. vamos nos concentrar em aprofundar o debate e a democracia direta | recomendação: democracia direta - Uma democracia direta é qualquer forma de organização na qual todos os cidadãos podem participar diretamente no processo de tomada de decisões. As primeiras democracias da antiguidade foram democracias diretas.

177 usuario3: a pec 71 seria interessante, a proposta é unificar as eleições nacionais e municipais | recomendação: PROPOSTA DE EMENDA À CONSTITUIÇÃO № 00071, de 2012 - Estende para seis anos o mandato dos Prefeitos, Vice-Prefeitos e Vereadores eleitos em 2016, para estabelecer a coincidência geral das eleições.

178 usuario11: então a pec 71 iria impedir que um político que se candidatasse para algo mudasse de cargo daqui a 2 anos... isso parece bom | recomendação: PROPOSTA DE EMENDA À CONSTITUIÇÃO Nº 00071, de 2012 - Estende para seis anos o mandato dos Prefeitos, Vice-Prefeitos e Vereadores eleitos em 2016, para estabelecer a coincidência geral das eleições.

179 usuario7: qualquer pessoa tem o direito de se candidatar. seja quem for. a consciência politica começa com o eleitor, ele deverá escolher em quem votar.

180 usuario8: mas nao teriamos ideiais em comum a seguir, bandeiras filosoficas que representassem a maioria do povo

181 usuario16: e temos massa crítica politizada o suficiente para mobilizações de monta? acho que não...

182 usuario3: não entendi usuario16

183 usuario15: e educação política se dá nas escolas, nas residências ... como fazer?

184 usuario6: eu acho que a proibição de coligações em eleições proporcionais uma medida essêncial. quando a coisa é majoriária, a aliança tende a ser programática | recomendação: coligações - Uma coligação ou coalizão política é um pacto entre dois ou mais partidos políticos, normalmente de ideias afins, para governar um país, uma região ou outra entidade administrativa. Na maioria dos regimes democráticos, as coligações são permitidas e originam-se quando um só partido ou grupo político não tem apoio suficiente na câmara legislativa correspondente (parlamento, senado ou ambos), como consequência dos resultados eleitorais.

185 usuario11: usuario6, concordo com a colocação, mas talvez seja importante para entendermos o cenário atual

186 usuario14: e regular que as coligações sejam duradouras, para diminuir o portunismo | recomendação: coligações - Uma coligação ou coalizão política é um pacto entre dois ou mais partidos políticos, normalmente de ideias afins, para governar um país, uma região ou outra entidade administrativa. Na maioria dos regimes democráticos, as coligações são permitidas e originam-se quando um só partido ou grupo político não tem apoio suficiente na câmara legislativa correspondente (parlamento, senado ou ambos), como consequência dos resultados eleitorais.

187 usuario11: mas se você coloca uma barreira em algo é natural que as pessoas queiram quebrá-la

188 usuario6: pode ser sim, uma vez que se colque em discussão um projeto policto. ser majoritaria não significa que sso vai atender a um grupo. é só uma questão de como se distribuirão as vagas

189 usuario6: o problema da clausula de barreira, é que ela pode favorecer os partidos sem rosto tipo o pmdb | recomendação: pmdb - Partido do Movimento Democrático Brasileiro (PMDB) é o maior partido político brasileiro, com 2 355 472 filiados (maio de 2012),6 apesar de não ter até hoje elegido nenhum Presidente da República através do voto direto. Fundado em 1980, possui uma orientação política

190 usuario6: as democracias europeias passaram pelo mesmo processo longo que nos passamos

191 usuario3: pessoal, acho que já podemos encerrar o chat. gostaria muito de saber o que vocês acharam do chat bem como as recomendações, críticas e sugestões

192 usuario5: gostei muito pessoal! acredito que no futuro a tecnologia nos possibilitará exercer a democracia direta? que afinal seria a democracia "original". vou responder o questionário! abraços | recomendação: democracia direta - Uma democracia direta é qualquer forma de organização na qual todos os cidadãos podem participar diretamente no processo de tomada de decisões.

193 usuario7: sempre haverá espaço para discussão e mobilização. revolta e insubordinação é coisa do passado. hoje vivemos em pleno estado democrático. isso é coisa do pig

194 usuario3: o que é pig usuario7?

195 usuario7: partido da imprensa golpista. | recomendação: partido da imprensa golpista - Partido da Imprensa Golpista (comumente abreviado para PIG ou PiG) é uma expressão usada por órgãos de imprensa e blogs políticos de orientação de esquerda para se referir a órgãos de imprensa e jornalistas por eles considerados tendenciosos, que se utilizariam do que chamam grande mídia como meio de propagar suas ideias e tentar desestabilizar governos de orientação política contrária

196 usuario6: não entendi usuario7?

197 usuario6: usuario7 quem vive o pelno estado democratico de direito? nós, ou o amarildo?

198 usuario10: isso é verdade usuario6

199 usuario6: pessoal tenho que ir

200 usuario6: abraços

201 usuario3: O questionario para avaliação está nesse link

202 usuario14: o link não está funcionando usuario3

203 usuario3: aqui funciona clicando

204 usuario13: copia o link e cola no navegador que vai

205 usuario13: o 1º link não está funcionando

206 usuario14: sim, clicando no final, funciona!

207 usuario14: pessoal! muito obrigada pela ajuda e participação! peço que não deixem de dar usa

opiniao e sugestoes no questionario

208 usuario3: obrigado pessoal, foi ótima a discussão

209 usuario16: boa noite!

210 usuario3: boa noite a todos211 usuario8: boa noite pessoal

212 usuario13: boa noite213 usuario10: boa noite.