

UNIVERSIDADE DE SÃO PAULO
Instituto de Ciências Matemáticas e de Computação

Uma Interface para Visualização de Sumários
Multidocumento

Jader Bruno Pereira Lima



Uma Interface para Visualização de Sumários Multidocumento

Jader Bruno Pereira Lima

Orientador: Thiago Alexandre Salgueiro Pardo

Monografia de conclusão de curso apresentada ao
Instituto de Ciências Matemáticas e de Computação –
ICMC-USP - para obtenção do título de Bacharel em
Ciências de Computação.

Área de Concentração: Processamento de Linguagem
Natural

USP – São Carlos
Novembro de 2013

Dedicatória

Dedico este trabalho, assim como todo o meu percurso de sucesso na universidade, aos meus pais que me apoiam desde sempre.

Resumo

Normalmente, um usuário que tem em mãos um sumário produzido automaticamente a partir de uma grande quantidade de textos (por exemplo, provenientes de resultados de busca em um buscador como o Google ou em um buscador especializado em notícias como o Google Notícias) deseja ser capaz de visualizar adequadamente este sumário e explorar funcionalidades tanto simples quanto avançadas, como alternar visualmente entre o sumário e os textos de origem correspondentes a uma determinada porção textual do sumário delimitada pelo usuário, procurar nos textos por informações complementares a alguma informação presente no sumário, comprimir ainda mais ou expandir a quantidade de informação do sumário de acordo com o tempo disponível e o interesse do usuário no tópico abordado, filtrar o conteúdo do sumário em função de alguma palavra-chave ou termo de interesse, colorir e comentar determinadas passagens do sumário, traduzir automaticamente o sumário para uma outra língua de interesse, e disponibilizar um link a outros usuários para acesso a esse sumário produzido, dentre várias outras. Neste trabalho, propõe-se a investigação de algumas destas funcionalidades e, com base nisso, desenvolve-se e avalia-se uma interface visual de interação com o usuário.

Sumário

LISTA DE FIGURAS.....	VI
LISTA DE TABELAS.....	VII
CAPÍTULO 1: INTRODUÇÃO.....	1
1.1. CONTEXTUALIZAÇÃO E MOTIVAÇÃO.....	1
1.2. OBJETIVOS.....	3
1.3. ORGANIZAÇÃO DA MONOGRAFIA.....	3
CAPÍTULO 2: REVISÃO BIBLIOGRÁFICA.....	4
2.1. CONSIDERAÇÕES INICIAIS.....	4
2.2. CONCEITOS E TÉCNICAS RELEVANTES.....	4
2.3. TRABALHOS RELACIONADOS.....	5
2.4. CONSIDERAÇÕES FINAIS.....	11
CAPÍTULO 3: DESENVOLVIMENTO DO TRABALHO.....	12
3.1. CONSIDERAÇÕES INICIAIS.....	12
3.2. ESTUDO DE FUNCIONALIDADES DE VISUALIZAÇÃO.....	12
3.3. DESENVOLVIMENTO DO SISTEMA VISUM.....	16
3.4. AVALIAÇÃO DE USABILIDADE DO SISTEMA VISUM.....	20
3.5. DIFICULDADES, LIMITAÇÕES E TRABALHOS FUTUROS.....	26
3.6. CONSIDERAÇÕES FINAIS.....	26
CAPÍTULO 4: CONCLUSÃO.....	27
4.1. CONTRIBUIÇÕES.....	27
4.2. CONSIDERAÇÕES SOBRE O CURSO DE GRADUAÇÃO.....	27
REFERÊNCIAS.....	29
APÊNDICE A – QUESTIONÁRIO DE AVALIAÇÃO DO SISTEMA VISUM.....	31

Lista de Figuras

Figura 1: Exemplo de Sumário Multidocumento	2
Figura 2: Sumário gerado pelo sistema Columbia Newsblaster.....	6
Figura 3: Sistema de SAM News in Essence	7
Figura 4: Sistema de SAM Ineats	8
Figura 5: Interface inicial do sistema SumView	9
Figura 6: Sumário gerado pelo sistema SumView	9
Figura 7: Interface inicial do sistema RSumm	10
Figura 8: Sumário gerado pelo sistema RSumm	10
Figura 9: Interface do sistema ViSUM.....	19

Lista de Tabelas

Tabela 1: Funcionalidades encontradas nos sistemas de SAM estudados.....	13
Tabela 2: Importância das funcionalidades de visualização estudadas	14
Tabela 3: Classificação das funcionalidades com relação aos grupos de usuários	15
Tabela 4: Avaliação da importância das funcionalidades do sistema ViSUM.....	23
Tabela 5: Avaliação da facilidade de uso das interações do sistema ViSUM.....	24
Tabela 6: Avaliação das características de layout do sistema ViSUM.....	25

CAPÍTULO 1: INTRODUÇÃO

1.1. Contextualização e Motivação

Aplicações computacionais capazes de tratar a grande quantidade de informação disponibilizada atualmente na web têm se tornado cada vez mais necessárias. Em um informe recente da *International Data Corporation* (IDC), mostra-se que, somente em 2012, a web deve ter sido responsável pela disponibilização de 2.7 *zettabytes* de dados¹, uma quantidade nove vezes maior do que a produzida cinco anos atrás. Neste contexto, temos um leque de possibilidades de desenvolvimento de pesquisas que tratem dessa enorme quantidade de informação, que são disponibilizadas principalmente de forma textual.

Muitas pessoas utilizam a internet como principal fonte de aquisição de conhecimento e notícias do cotidiano, e é desejável que esta tarefa seja rápida e prática. Na área de Processamento de Linguagem Natural (PLN), uma subárea de Inteligência Artificial, há uma grande linha de pesquisa voltada em auxiliar na questão de aquisição de conhecimento de caráter textual e no tratamento destes dados de forma a torná-los mais compreensíveis para os usuários: a Sumarização Automática (SA). A tarefa de SA consiste na produção de uma versão mais curta de um texto fonte (Mani, 2001), que é chamado de sumário, ou resumo. Este sumário deve conter as informações mais importantes do texto fonte. Nesta área de SA, existe uma vertente de pesquisa que visa a criação de sumários a partir de vários textos que versam sobre um mesmo assunto, ou tópico: a Sumarização Automática Multidocumento (SAM) (Mckeown e Radev, 1995; Radev e Mckeown, 1998).

O resultado esperado de ambos os processos descritos acima, a SA e SAM, é o sumário textual, o qual provê para o usuário uma forma de adquirir o conhecimento de maneira rápida e resumida. Porém, quando focamos na SAM, é possível que sejam disponibilizadas mais informações sobre os dados sumarizados, além de apenas o sumário

¹ <http://www.idc.com/getdoc.jsp?containerId=prUS23177411#.UUnEZVvwLDM>

em si. Como o sumário é criado a partir de fragmentos extraídos de outros textos fonte, seria interessante se fosse possibilitado ao usuário o rastreamento desses textos fonte a partir de cada fragmento destes contidos no sumário, de forma a enriquecer a experiência do usuário na tarefa de leitura. Além desta funcionalidade de rastreamento, outras funcionalidades podem ser implementadas em um sistema de SAM, como a exibição de imagens ilustrativas do sumário, que podem adicionar informações relevantes para o seu entendimento; a exibição de palavras chave do sumário, que podem tornar mais prática e rápida a contextualização do tema do sumário para o usuário; ou a parametrização do tamanho do sumário de acordo com o tempo disponível e o interesse do usuário no tópico abordado. Estas, e outras funcionalidades, são discutidas neste trabalho.

A seguir, na Figura 1, temos um sumário automático gerado pelo sistema CSTSumm (Castro Jorge e Pardo, 2011) a partir de três textos jornalísticos que versam sobre um mesmo assunto. Nota-se que seria interessante se, além da visualização do sumário, fossem disponibilizadas algumas das funcionalidades citadas acima, como a visualização das notícias originais, para que o usuário tivesse a opção de aprofundar o seu nível de informação sobre aquele assunto.

Uma nova série de ataques criminosos foi registrada na madrugada desta segunda-feira, dia 7, em São Paulo e municípios do interior paulista. Os bandidos atacaram agências bancárias, bases policiais e prédios públicos com bombas e tiros. As ações são atribuídas à facção criminosa Primeiro Comando da Capital (PCC), que já comandou outros ataques em duas ocasiões. Eles tinham prometido retomar os ataques no Estado de São Paulo no Dia dos Pais, no próximo domingo. A promessa aparentemente começou a ser cumprida na madrugada de hoje. Cidades do interior, como Jundiaí, foram alvo de ataques. Na região do ABC Paulista, pelo menos dez ônibus foram incendiados - sete em Mauá e três em Santo André. Na capital, houve ataques a outros quatro ônibus. Uma bomba caseira foi jogada contra o prédio do Ministério Público, na capital do estado. A Secretaria da Fazenda também foi atingida por uma bomba. Duas bases da Guarda Civil Metropolitana (GCM), sendo uma no Capão Redondo, Zona Sul de São Paulo, foram alvo dos criminosos.

Figura 1: Exemplo de Sumário Multidocumento

1.2. Objetivos

Neste trabalho, tivemos como proposta a análise e a avaliação de funcionalidades de visualização presentes em sistemas existentes de SAM, com o objetivo de confirmar a hipótese inicial de que é possível melhorar a experiência do usuário na tarefa de leitura de sumários multidocumento, aumentando e aprofundando a sua aquisição de conhecimento. Além disso, propôs-se a construção de um sistema computacional de visualização de sumários multidocumento, chamado ViSUM – Sistema de Visualização de Sumários Multidocumento, contendo as principais funcionalidades estudadas.

1.3. Organização da Monografia

A seguir, na Seção 2, apresenta-se um relato sobre a revisão bibliográfica realizada neste projeto. Nesta revisão encontram-se os sistemas de SAM analisados e também um estudo sobre os tipos de usuários de sistemas de SA. Na Seção 3, são relatadas as atividades desenvolvidas neste trabalho, as quais englobam o estudo das funcionalidades encontradas nos sistemas de SAM apresentados na Seção 2 e o desenvolvimento e a avaliação do sistema computacional proposto, com base nesta análise.

CAPÍTULO 2: REVISÃO BIBLIOGRÁFICA

2.1. Considerações Iniciais

A seguir, na Seção 2.2, apresenta-se um relato do estudo que serviu como base para o desenvolvimento das atividades propostas neste projeto. Na Seção 2.3 são descritos alguns dos principais sistemas de sumarização multidocumento presentes na literatura.

2.2. Conceitos e Técnicas Relevantes

Uma forma de se estudar as funcionalidades presentes nos sistemas de SAM é focar no entendimento dos usuários típicos destes sistemas. Em uma visão generalista, os usuários podem ser divididos em dois grupos, dadas as suas necessidades de interação com o sumário (Mani e Maybury, 1999).

O primeiro grupo de usuários, chamado de “usuários generalistas”, necessita de uma visualização mais ampla, a qual quantifica os dados em vez de detalhá-los. Esses usuários necessitam de informações estatísticas e classificatórias sobre os dados. A intenção das funcionalidades voltadas para este grupo de usuários é a organização e disponibilização de dados que tornem a experiência de leitura mais prática e que adicionem informações que possam direcionar a leitura do sumário, de modo que a torne mais rápida. Um exemplo de uma funcionalidade voltada para este grupo de usuários é a exibição das palavras chave do sumário, as quais fornecem rapidamente os principais tópicos abordados, agilizando o processo de ganho de informação. Outro exemplo é a exibição do número de acessos de cada texto fonte na web, que forneceria a informação de onde começar a leitura, partindo do princípio de que as textos mais acessados são mais difundidos e importantes para aquele assunto.

Entretanto, existe um segundo grupo de usuários, chamado de “usuários especialistas”, o qual necessita de uma abordagem mais detalhista, que engloba técnicas de visualização que possibilitem a ampliação da quantidade de dados a serem exibidos. A intenção principal das funcionalidades voltadas para este grupo de usuários é maximizar a

aquisição de informação, oferecendo diversos níveis de profundidade de exibição dos dados. Um exemplo de uma funcionalidade voltada para este grupo de usuários é a parametrização do sumário, que possibilita ao usuário a criação de um texto com o nível de detalhes que desejar, expandindo assim a aquisição de conhecimento. Outra funcionalidade voltada para este grupo é a possibilidade de navegação entre os textos fonte, possibilitando ao usuário a extração de conhecimento utilizando todos os dados disponíveis no sistema de SAM.

2.3. Trabalhos Relacionados

Neste trabalho, foram estudados vários sistemas de sumarização multidocumento, com o foco principalmente em sistemas de SAM voltados para a sumarização de notícias. Os sistemas estudados foram *Columbia Newsblaster* (McKeown et al., 2002), *News in Essence* (Radev et al., 2004), *Ineats* (Leuski et al., 2000), *RSumm* (Ribaldo et al., 2012) e *SumView* (Wang et al., 2012). Neste estudo, pôde-se verificar e analisar a presença de várias funcionalidades de visualização de sumários.

A seguir, na Figura 2, temos a visualização de um sumário gerado automaticamente pelo sistema *Columbia Newsblaster* (McKeown et al., 2002). Este sistema de SAM online, com foco em notícias, é considerado um dos melhores na literatura e, por esse motivo, é também um dos mais difundidos e utilizados. Na figura, são indicadas com números algumas das funcionalidades existentes neste sistema, descritas a seguir: (1) direcionamento para o texto fonte, de onde foi retirada cada sentença presente no sumário; (2) dados dos textos fonte, como data de publicação, endereço na web e número de páginas; (3) palavras-chave referentes aos textos sumarizados; (4) imagens ilustrativas sobre os tópicos abordados no sumário; e (5) outros sumários sugeridos, a partir do texto sumarizado pelo usuário.

Columbia Newsblaster
Summarizing all the news on the Web

Thursday, February 16, 2012
Articles from 02/13/2012 to 02/16/2012
Last update: 4:25 AM EST

Assad forces regain Damascus suburbs, say Syrian activists

Summary from multiple countries, from articles in English
[UPDATED] (see summary with new information since yesterday)

Troops seized eastern suburbs of Damascus from rebels late on Sunday, opposition activists said, after two days of fighting only a few kilometers from the centre of power of Syrian President Bashar Assad. (article 1) Late last month, a UN commission of inquiry on Syria said that Syrian military and security forces had committed crimes against humanity including murder, torture and rape and that Assad's government of bore responsibility. (article 5) As Syrian forces stepped up their assault Wednesday on rebellious cities, President Bashar al-Assad ordered a referendum on a new constitution that would create a multiparty system in a country that has been ruled by his autocratic family dynasty for 40 years. (article 2) After 11 months of bloodshed, with well over 5,000 dead in the regime's crackdown on protesters and rebels, Assad's opponents say the referendum and reforms are not enough and that the country's strongman must go. (article 3) Assad on Wednesday approved a draft for a new constitution that would end his ruling Baath party's monopoly on power in the fraught nation and open the door for all political parties to compete. (article 4) VIENNA - Russia's foreign minister is blaming Western U.N. Security Council members for the failure to reach agreement on a resolution about Syria. (article 7)

Other summaries about this story:

- Summary from Canada, from articles in English (1 articles) [compare]
- Summary from Israel, from articles in English (0 articles) [compare]
- Summary from United States, from articles in English (4 articles) [compare]

Other stories about Assad, Hamas and Syria:

- Iran threatening to cut Hamas funds, arms supply if it flees Syria (4 articles)

Event tracking:

- Track this story's development in time

Story keywords
Assad, Hamas, Syria, Syrian, Damascus

Source articles

1. Assad forces regain Damascus suburbs, say Syrian activists (haaretz.com, 02/16/2012, 892 words)
2. Syria's Assad calls for national vote, but steps up assault on uprising (Washington Post, 02/16/2012, 408 words)
3. Syria's Assad calls for vote but steps up assault (cbc.ca, 02/15/2012, 965 words)

Figura 2: Sumário gerado pelo sistema Columbia Newsblaster

News in Essence (Radev et al., 2004), assim como o *Columbia Newsblaster*, também é um sistema de SAM online, focado na sumarização de notícias. Uma de suas principais características é a possibilidade da escolha do conjunto de textos a ser sumarizados. Neste sistema, notam-se algumas funcionalidades interessantes, como a referência aos textos fonte a partir da sentença escolhida presente no sumário; a possibilidade de escolha do tamanho do sumário e a visualização de informações sobre os textos fonte, por exemplo, a data e a hora de publicação. Um exemplo da visualização de um sumário no sistema *News in Essence* é apresentado a seguir, na Figura 3. Nota-se que este sistema é voltado para usuários que necessitam de informações qualitativas dos dados sumarizados, pois contém funcionalidades que permitem o aumento da quantidade de informação a ser visualizada pelo usuário.

...www...NewsInEssence...com...

Interactive Multi-source News Summarization

Home

[Current Clusters](#)

[Create Cluster](#)

[Summarize Cluster](#)

[Track Cluster](#)

[User Cluster Archive](#)

[CIDR Cluster Archive](#)

[Google Cluster Archive](#)

Help

[About NewsInEssence](#)

[Contact Us](#)

CLAIR

MEAD

[summarization.com](#)

Ford reports 1.2 billion loss in first quarter

Ford reports 1.2 billion 1Q loss, worst performance in four years. Ford reports \$1.2 billion loss in first quarter took a massive restructuring charge in the first quarter, pushing its quarterly loss to \$1.2 billion, down \$2.4 billion from a year earlier.

[\[4 Articles from 4 Sources\]](#) [\[4 Summaries\]](#)

Recent User Clusters (more)

- ['China s president looks to soothe an uneasy U.S. public during visit'](#)
14 articles, 3 summaries: 04/20, 5:48 PM
- ['Iran s president warns his army will cut the hand of any aggressor'](#)
2 articles, 3 summaries: 04/19, 12:38 AM
- ['China eases foreign currency rules ahead of Hu s U.S. trip'](#)
10 articles, 3 summaries: 04/17, 4:52 PM

Recent CIDR Clusters (more)

- ['Pomp, embarrassment greet Hu on first visit to White House as China's leader'](#)
43 articles, 4 summaries: 04/21, 11:44 AM
- ['Giant flag at Windsor Castle heralds 80th birthday of Queen Elizabeth II'](#)
10 articles, 4 summaries: 04/21, 11:44 AM
- ['Iraqi interim PM tells party he'd step down Apr 20, 2006'](#)
8 articles, 4 summaries: 04/21, 11:44 AM

[NIE Headlines](#)
[Build your own cluster of articles.](#)

NewsTroll from URL:
To search using default values, click "Search", to modify settings, use the [form below](#)

NewsTroll from query:

[Advanced Options](#)

User Clusters (Archive)

- ['China s president looks to soothe an uneasy U.S. public during visit'](#)
14 articles, 3 summaries: 04/20, 5:48 PM
- ['Iran s president warns his army will cut the hand of any aggressor'](#)
2 articles, 3 summaries: 04/19, 12:38 AM
- ['China eases foreign currency rules ahead of Hu s U.S. trip'](#)
10 articles, 3 summaries: 04/17, 4:52 PM
- ['Middle East Rough start for Palestinian parliament'](#)
15 articles, 3 summaries: 04/17, 1:32 AM
- ['Canada's best source for news continuously updated from The Globe and Mail'](#)
1 article, 3 summaries: 04/15, 6:45 PM
- ['Help Your guide to the Player'](#)
12 articles, 3 summaries: 04/14, 9:03 AM

Ford reports 1.2 billion loss in first quarter

produced on 04/21, 11:44 AM

2% Summary

Ford reports 1.2 billion 1Q loss, worst performance in four years (2:1)
Ford reports \$1.2 billion loss in first quarter took a massive restructuring charge in the first quarter, pushing its quarterly loss to \$1.2 billion, down \$2.4 billion from a year earlier. (4:3)

Summaries of all documents: [\[2%\]](#) [\[5%\]](#) [\[10%\]](#) [\[20%\]](#)

Cluster Documents

- 1. [Business US problems fuel huge Ford loss](#) [\[Cached\]](#) [\[news.bbc.co.uk\]](#) [\[Use As Seed\]](#)
- 2. [Ford reports 1.2 billion 1Q loss, worst performance in four years](#) [\[Cached\]](#)

Figura 3: Sistema de SAM News in Essence

O sistema *Ineats* (Leuski et al., 2000) nos possibilita outras funcionalidades de visualização, como a utilização de cores para referenciar os textos fonte das sentenças escolhidas para compor o sumário; a visualização de um mapa, contendo os locais de publicação dos textos fonte; a possibilidade de parametrização do sumário com relação ao número de palavras, número de sentenças e porcentagem de informação redundante entre as sentenças e a escolha dos tópicos mais importantes para a sumarização. Nos artigos estudados sobre este sistema, também não foi encontrada uma avaliação que considerasse a interface como parâmetro. Um exemplo de um sumário visualizado neste sistema é apresentado na Figura 4, a seguir. Nota-se que este sistema é voltado para usuários que necessitam de informações quantitativas sobre os dados apresentados pelo usuário, pois contém funcionalidades que auxiliam na visualização de informações que explicam os dados em vez de detalha-los.

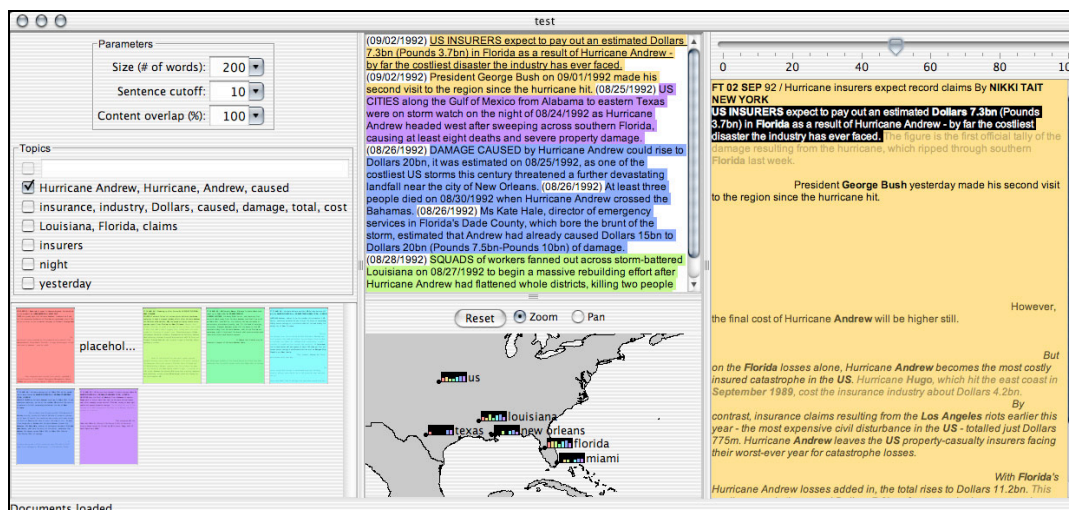


Figura 4: Sistema de SAM Ineats

SumView (Wang et al., 2012) é um sistema produzido no laboratório em que o aluno desenvolveu suas atividades do projeto BEPE/FAPESP, durante seu trabalho de conclusão de curso. Trata-se de um sistema de sumarização que produz um sumário contendo os comentários mais relevantes sobre um determinado produto à venda, os quais são separados por determinadas características, ou aspectos. Nas Figuras 6 e 7 temos um exemplo de execução do sistema *SumView*.

Na Figura 6, é disponibilizada ao usuário a opção de escolher entre algumas marcas de chaleiras automáticas. Para a produção do sumário, o sistema fornece alguns aspectos que possam ser relevantes para aquele tipo de produto. Neste exemplo, os aspectos sugeridos são *boil* (fervura), *cord* (fiação), *price* (preço), *problem* (problemas) e *lid* (tampa). Além desses aspectos, o usuário pode inserir algum outro aspecto que julgar interessante para aquele tipo de produto. Na Figura 7, é apresentado ao usuário o sumário do produto escolhido, para cada um dos aspectos solicitados. Deste sistema, pode-se retirar uma funcionalidade interessante para a visualização de sumários multidocumento: a separação da visualização do conteúdo em subpartes, contendo os subtópicos do assunto em questão, a qual possibilita um direcionamento melhor aos interesses do usuário. Neste caso do sistema *SumView*, os subtópicos seriam os aspectos e o assunto a ser sumarizado seriam as opiniões sobre um produto a venda. Assim como nos sistemas analisados

anteriormente, nos artigos estudados sobre este sistema também não foi encontrada uma avaliação que considerasse a interface como parâmetro.



Figura 5: Interface inicial do sistema SumView



Figura 6: Sumário gerado pelo sistema SumView

A seguir, nas Figuras 7 e 8, apresenta-se mais um sistema estudado neste projeto: o sistema *RSumm* (Ribaldo et al., 2012). Este sistema de SAM, desenvolvido no NILC – Núcleo

Interinstitucional de Linguística Computacional (o mesmo laboratório onde foram desenvolvidas as atividades deste projeto), produz um sumário a partir de textos jornalísticos recuperados do site Google Notícias que versam sobre um mesmo assunto. Deste sistema podemos retirar as seguintes funcionalidades de visualização: referência em cada sentença escolhida para compor o sumário ao seu respectivo texto fonte; limpeza das sentenças no sumário (como remoção de *tags HTML*, por exemplo); textos fonte retirados de uma única API, no caso o Google Notícias; e parametrização do tamanho do sumário.



Figura 7: Interface inicial do sistema RSumm

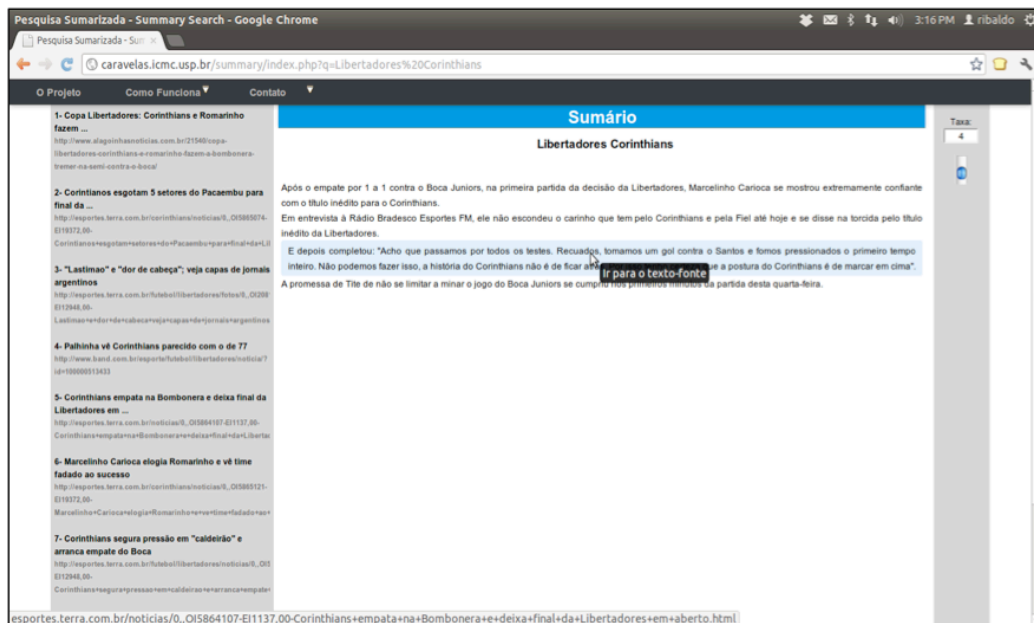


Figura 8: Sumário gerado pelo sistema RSumm

Além destes trabalhos, outros sistemas relacionados de visualização foram estudados nesta revisão bibliográfica. O sistema Google Trends² possui funcionalidades de visualização que se adequam a usuários que buscam informações mais quantitativas sobre um determinado assunto, por exemplo, a evolução histórica do número de buscas deste assunto no site de buscas ou a localização geográfica dos dispositivos de onde são requisitadas as buscas. O sistema PEx – *Projection Explorer* (Paulovich, 2007), desenvolvido no ICMC, é uma ferramenta que pode ser utilizada para exploração de representações visuais de coleções de documentos textuais, auxiliando o usuário a entender o seu conteúdo. Os sistemas Sensebot³ e Opimining⁴ (este último desenvolvido no NILC) possuem a funcionalidade de identificação da quantidade de opiniões positivas e negativas sobre algum termo pesquisado pelo usuário, presente nos textos fonte recuperados de um repositório online.

2.4. Considerações Finais

A partir do levantamento das funcionalidades encontradas nos sistemas apresentados na seção anterior, foi realizado um estudo sobre a importância dessas funcionalidades na legibilidade do sumário. Este estudo e a descrição do sistema proposto estão relatados na Seção 3, a seguir.

² <http://www.google.com.br/trends>

³ <http://www.sensebot.net/>

⁴ <http://opimining.appspot.com/>

CAPÍTULO 3: DESENVOLVIMENTO DO TRABALHO

3.1. Considerações Iniciais

As funcionalidades catalogadas na Seção 2.3 foram analisadas e agrupadas de acordo com as suas características. Esta análise se encontra na Seção 3.2. Na Seção 3.3, temos a descrição do sistema ViSUM, que foi projetado e desenvolvido neste trabalho. As funcionalidades presentes neste sistema foram escolhidas com base no estudo realizado sobre as funcionalidades encontradas nos sistemas estudados na revisão bibliográfica. A Seção 3.4 relata a avaliação de usabilidade do sistema ViSUM, que foi feita com usuários reais.

3.2. Estudo de Funcionalidades de Visualização

As funcionalidades encontradas nos sistemas de SAM analisados foram divididas em três grupos. O primeiro contém as funcionalidades que tratam do modo como a informação é exibida ao usuário. O segundo grupo contém as funcionalidades que dizem respeito aos métodos utilizados para enriquecer a interação do usuário com as informações contidas no sumário. Por fim, o terceiro grupo contém as funcionalidades que tratam da parametrização do sistema de sumarização pelo usuário. A seguir, na Tabela 1, estão listadas todas as funcionalidades encontradas nos sistemas que foram possíveis ser testados: *Columbia Newsblaster* (Mckeown et al., 2002), *News in Essence* (Radev et al., 2004), *Ineats* (Leuski et al., 2000), *RSumm* (Ribaldo et al., 2012) e *SumView* (Wang et al., 2012). Na tabela, indica-se, para cada sistema, se a funcionalidade estava presente ou não, por exemplo, nota-se que a funcionalidade 1.1 “Referência aos textos-fonte” faz parte de todos os sistemas, exceto do sistema *SumView*.

Tabela 1: Funcionalidades encontradas nos sistemas de SAM estudados

	Columbia Newsblaster	News in Essence	Ineats	RSumm	SumView
Grupo 1: Visualização de Informação					
1.1: Referência aos textos fonte	Sim	Sim	Sim	Sim	Não
1.2: Destaque da sentença escolhida em seu texto-fonte	Sim	Sim	Não	Não	Não
1.3: Adequação das informações textuais (remoção de tags HTML, por exemplo)	Sim	Não	Sim	Sim	Sim
1.4: Exibição das palavras chave do sumário	Sim	Não	Sim	Não	Sim
1.5: Exibição da quantidade de palavras dos textos-fonte	Sim	Não	Sim	Não	Sim
1.6: Exibição da data de publicação dos textos-fonte	Sim	Não	Não	Não	Não
1.7: Exibição do local de publicação dos textos-fonte	Não	Não	Sim	Não	Não
Grupo 2: Funcionalidades avançadas de visualização					
2.1: Assuntos mais sumarizados pelos usuários	Não	Sim	Não	Não	Não
2.2: Exibição de todos os textos-fonte utilizados	Não	Sim	Não	Sim	Não
2.3: Exibição de imagens ilustrativas do sumário	Sim	Não	Não	Não	Sim
2.4: Diferenciação das sentenças por cores de acordo com seu tópico ou texto-fonte	Não	Não	Sim	Não	Não
2.5: Agrupamento dos textos-fonte por tópicos	Sim	Não	Não	Não	Sim
Grupo 3: Parametrização do sumarizador					
3.1: Parametrização do tamanho do sumário	Não	Sim	Sim	Sim	Não
3.2: Escolha dos tópicos mais relevantes a serem sumarizados	Não	Não	Sim	Não	Sim

A partir deste levantamento, pôde-se inferir alguns dados sobre este cenário de sistemas de SAM na literatura, com foco em suas funcionalidades. Quanto mais comum é a presença de uma funcionalidade nos sistemas estudados, pode-se dizer que esta funcionalidade é mais importante para o processo de leitura do sumário multidocumento. A seguir, na Tabela 2, temos a classificação destas funcionalidades referente à sua importância em um sistema de visualização de sumário multidocumento. Esta tabela foi construída seguindo o seguinte critério: as funcionalidades presentes em pelo menos 80% dos sistemas analisados foram consideradas como indispensáveis em um sistema com esta finalidade. As funcionalidades presentes entre 50% e 79% dos sistemas analisados foram consideradas como importantes. As demais funcionalidades levantadas foram consideradas como pouco importantes. O número de asteriscos, na tabela, é proporcional à indispensabilidade da funcionalidade, sendo que *** representa as funcionalidades indispensáveis, e * representa as funcionalidades pouco importantes.

Tabela 2: Importância das funcionalidades de visualização estudadas

Funcionalidades Indispensáveis ***
1.1: Referência aos textos fonte ***
1.3: Adequação das informações textuais (remoção de <i>tags</i> HTML, por exemplo) ***
Funcionalidades Importantes **
1.4: Exibição das palavras chave do sumário **
1.6: Exibição da data de publicação dos textos-fonte **
3.1: Parametrização do tamanho do sumário **
Funcionalidades Pouco Importantes **
1.2: Destaque da sentença escolhida em seu texto-fonte *
1.5: Exibição da quantidade de palavras dos textos-fonte *
1.7: Exibição do local de publicação dos textos-fonte *
2.1: Assuntos mais sumarizados pelos usuários *
2.2: Exibição de todos os textos-fonte utilizados *
2.3: Exibição de imagens ilustrativas do sumário *
2.4: Diferenciação das sentenças por cores de acordo com seu tópico ou texto-fonte *
2.5: Agrupamento dos textos-fonte por tópicos *
3.2: Escolha dos tópicos mais relevantes a serem sumarizados *

Além desta classificação, as funcionalidades podem ser divididas com relação ao tipo de usuário que é seu foco. Mani e Maybury definiram dois grupos de usuários de SAM, os quais estão descritos na Seção 1 deste relatório, que são basicamente grupos de usuários que necessitam de ferramentas que quantifiquem e qualifiquem os dados do sumário. A Tabela 3, a seguir, nos mostra a divisão destas funcionalidades nestes dois grupos de usuários. Esta divisão foi feita da seguinte forma: todas as funcionalidades que expandem a aquisição de conhecimento do usuário, além da simples leitura do sumário, foram consideradas como funcionalidades qualitativas. Por exemplo, a funcionalidade “1.1: Referência aos textos fonte” adiciona informação ao sistema, além do sumário apresentado ao usuário, e por este motivo é classificada como qualitativa. Todas as funcionalidades que resumem ou classificam os dados já contidos no sumário, foram consideradas como funcionalidades quantitativas. Por exemplo, a funcionalidade “Exibição de palavras chave do sumário” tem o objetivo de facilitar e resumir a leitura do usuário por meio de palavras chave. Note que esta informação já é contida no sumário em si, porém esta funcionalidade facilita e agiliza a aquisição de informação do usuário e, por esse motivo, é classificada como quantitativa.

Tabela 3: Classificação das funcionalidades com relação aos grupos de usuários

	Columbia Newsblaster	News in Essence	Ineats	RSumm	SumView
Funcionalidades qualitativas					
1.1: Referência aos textos fonte ***	Sim	Sim	Sim	Sim	Não
1.2: Destaque da sentença escolhida em seu texto-fonte *	Sim	Sim	Não	Não	Não
1.3: Adequação das informações textuais (remoção de tags HTML, por exemplo) ***	Sim	Não	Sim	Sim	Sim
1.6: Exibição da data de publicação dos textos-fonte **	Sim	Não	Não	Não	Não
1.7: Exibição do local de publicação dos textos-fonte *	Não	Não	Sim	Não	Não
2.2: Exibição de todos os textos-fonte utilizados *	Não	Sim	Não	Sim	Não
2.3: Exibição de imagens ilustrativas do sumário *	Sim	Não	Não	Não	Sim
3.1: Parametrização do tamanho do sumário **	Não	Sim	Sim	Sim	Não
3.2: Escolha dos tópicos mais relevantes a serem sumarizados *	Não	Não	Sim	Não	Sim
Funcionalidades quantitativas					
1.4: Exibição das palavras chave do sumário **	Sim	Não	Sim	Não	Sim
1.5: Exibição da quantidade de palavras dos textos-fonte *	Sim	Não	Sim	Não	Sim
2.1: Assuntos mais sumarizados pelos usuários *	Não	Sim	Não	Não	Não
2.4: Diferenciação das sentenças por cores de acordo com seu tópico ou texto-fonte*	Não	Não	Sim	Não	Não
2.5: Agrupamento dos textos-fonte por tópicos *	Sim	Não	Não	Não	Sim

Analisando-se as Tabelas 2 e 3, nota-se que a os sistemas de SAM analisados seguem uma tendência de foco em usuários com necessidade de uma visualização mais qualitativa dos dados, que buscam a expansão e a otimização da quantidade de dados a ser apresentada ao mesmo tempo. Essas funcionalidades visam aumentar o alcance da aquisição de informação do usuário, possibilitando técnicas que facilitem e organizem os dados do sistema, que englobam não somente o sumário, mas também informações extras, por exemplo, os textos fonte. Com esta análise, podemos concluir que, em um sistema de SAM, a visualização de dados extras é tão importante quanto a visualização do sumário em si, e vemos que é necessário a implementação dessas funcionalidades para aumentar a experiência final do usuário.

Sobre alguns dos sistemas estudados, também podemos retirar algumas conclusões. Neste estudo, constatou-se que o *Columbia Newsblaster* é um dos melhores sistemas de visualização de SAM disponíveis atualmente, pois contém 8 das 14 funcionalidades levantadas, sendo que 6 destas são da categoria de “Visualização de Informação”. Conclui-se que este sistema atende muitas necessidades de iteração do usuário na tarefa de visualização e, por esse motivo, é tão difundido e utilizado.

O sistema *Ineats* também possui uma quantidade bastante satisfatória de funcionalidades, contendo 8 das 14 funcionalidades estudadas. Este sistema se sobressai em relação aos outros, pois é o único que implementa as duas funcionalidades de parametrização do sumarizador, possibilitando ao usuário um melhor controle do sumário final. Nota-se que, diferentemente do *Columbia Newsblaster*, esse sistema se baseia em otimizar o sumário em vez de prover novas técnicas de visualização de informações extras, como a visualização de palavras chave do sumário.

O sistema *News in Essence* é voltado para uma interface que visa agregar informação de vários usuários que utilizam o sistema, pois é o único sistema que implementa a funcionalidade de visualização dos assuntos mais sumarizados por usuários do sistema. Essa informação pode ser relevante, pois nos traz uma informação extra muito importante sobre os dados do sistema: o quão interessante é determinado assunto naquele momento para os usuários.

3.3. Desenvolvimento do Sistema ViSUM

Após o estudo de funcionalidades de visualização presentes em SAM, relatado nas seções anteriores, o aluno reuniu todas as características necessárias para um sistema com este propósito. Com isto, foi proposto o desenvolvimento de um sistema de visualização de sumários multidocumento. Este protótipo, chamado ViSUM – Visualizador de Sumários Multidocumento, foi projetado para ser um sistema web, onde pode ser executado por navegadores de internet, como o Internet Explorer. As tecnologias utilizadas foram a linguagem PHP para a programação no lado do servidor da aplicação, integrada com a linguagem de script JavaScript, responsável pelo tratamento de eventos no lado do cliente, além de tecnologias já difundidas de programação web, como o HTML e as folhas de estilo CSS.

O sistema ViSUM possui dois módulos: no primeiro, o sistema deve receber como dados de entrada todas as informações do processo de SAM, ou seja, o sumário gerado por meio de algum algoritmo de sumarização e os textos fonte que foram sumarizados por este algoritmo. A partir destes dados, cabe ao ViSUM fornecer as funcionalidades de visualização para o usuário final, que é o indivíduo que tem o interesse pelo assunto que

foi sumarizado. O segundo módulo consiste na recuperação automática dos textos fonte, a partir de uma busca por algum tópico realizada pelo usuário final. Neste caso, o processamento seria dividido em duas etapas. Na primeira etapa, o usuário final forneceria o tópico a ser sumarizado e, então, o sistema recuperaria os textos fonte de algum repositório de textos específico, neste caso, o Google Notícias. Na segunda etapa, o sistema envia esses textos fonte para algum sistema de sumarização, o qual é o responsável pela aplicação do algoritmo que selecionará as sentenças do sumário final.

Uma ressalva deve ser feita sobre este sistema: foi desenvolvido um sistema de visualização de SAM genérico, ou seja, que independe do algoritmo de escolha das sentenças do sumário. Basicamente, a ideia desse sistema ViSUM é ser um sistema que possa ser utilizado por dois tipos de usuário: o pesquisador que trabalha na área de pesquisa de SAM e deseja visualizar os resultados de seu sistema de maneira clara e utilizando funcionalidades de visualização; e o usuário comum, que deseja apenas a produção de um sumário a partir de um conjunto de textos que versam sobre um mesmo assunto, sem se importar com o algoritmo utilizado na SAM, e que busca apenas se informar sobre assuntos de seu interesse.

Para este sistema, foi escolhida uma abordagem focada em usuários que desejam uma visualização mais qualitativa dos dados, os quais necessitam de técnicas que aumentem e organizem a quantidade de informação a ser visualizada. A seguir, temos a descrição das funcionalidades presentes no sistema ViSUM, além da motivação para a sua ocorrência e de sua classificação com base em qual grupo de usuários é focada.

- F1: Visualização de todos os textos fonte (qualitativa). Vemos que é de suma importância a possibilidade de visualização de todos os textos fonte utilizados para a sumarização, pois é necessário prover ao usuário diferentes níveis de profundidade na aquisição do conhecimento.
- F2: Referência aos textos fonte, a partir das sentenças do sumário (qualitativa). Esta funcionalidade facilita e agiliza a interação do usuário com o sistema, pois, é facilitada a tarefa de rastreamento e identificação da sentença escolhida em seu texto original. Sua ação é disparada quando o usuário clica sobre a sentença no sumário.

- F3: Parametrização do tamanho do sumário (qualitativa). Pelo mesmo princípio de que é interessante prover diferentes níveis de profundidade de aquisição de conhecimento ao usuário na visualização do sumário, temos que também é interessante prover essas opções em um nível de produção do sumário, neste caso, possibilitando a parametrização do tamanho do sumário.
- F4: Exibição de imagens ilustrativas do sumário (qualitativa). A utilização de imagens é bastante utilizada por vários meios de comunicação para a ilustração do assunto apresentado a seus usuários, pois embute uma grande quantidade de informação.
- F5: Exibição de informações extras sobre o sumário gerado (quantitativa). Esta é a única funcionalidade quantitativa do sistema, a qual fornece dados sobre o número de palavras do sumário e as palavras mais frequentes, com a remoção de *stopwords*. Sua função é prover uma visualização mais ágil dos dados presentes no sumário ao usuário. Esta funcionalidade está presente no protótipo, pois sua ocorrência nos sistemas da literatura é bastante comum, como podemos ver na Seção 3.2.
- F6: Possibilidade de impressão do sumário e de seus textos fonte (qualitativa). É interessante possibilitar ao usuário a troca de mídia em que os dados são visualizados, neste caso, da tela eletrônica para o papel.
- F7: Adequação das informações textuais (por exemplo, remoção de *tags* HTML) (qualitativa). Como estamos lidando com mídias digitais, as informações necessitam de codificação para serem estruturadas. Vemos que se faz necessário o tratamento das informações a serem visualizadas para que estas informações de codificação não atrapalhem o usuário em sua leitura. No Sistema ViSUM esta remoção do conteúdo inadequado das páginas web é feita utilizando o sistema NCleaner (Evert, 2008).

Nota-se que nem todas as funcionalidades estudadas anteriormente foram escolhidas para a implementação, pois algumas foram consideradas de pouca ajuda no processo de aquisição de conhecimento pelo usuário. Todas as funcionalidades consideradas indispensáveis ou muito importantes (vide Tabela 2) foram projetadas neste protótipo, além de algumas das funcionalidades consideradas como pouco importantes,

porém que agregam bastante informação ao sistema. A seguir, na Figura 9, temos a interface do sistema ViSUM.

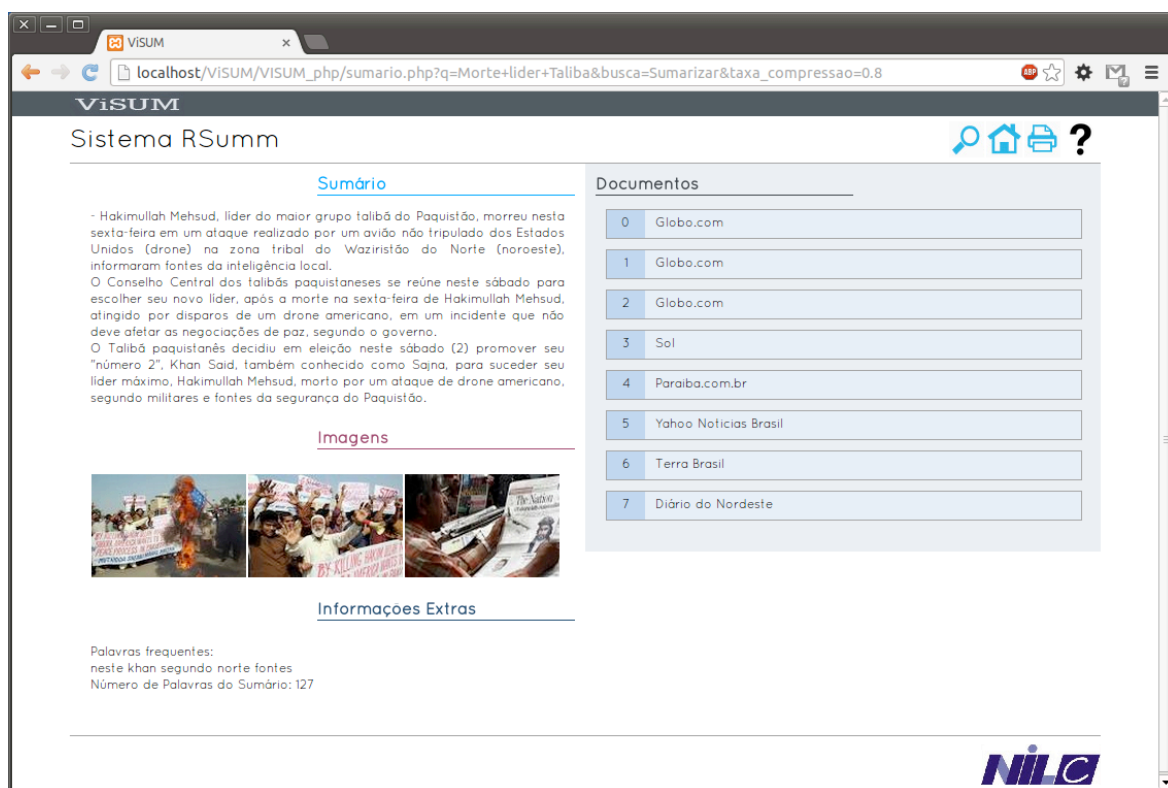


Figura 9: Interface do sistema ViSUM

Na Figura 9, podemos identificar todas as 7 funcionalidades descritas anteriormente. O sistema pode ser dividido em 4 áreas informativas: cabeçalho, área de visualização do sumário, área de visualização dos textos fonte e rodapé. No cabeçalho, estão contidos o logotipo do sistema, o nome do sistema que implementa o algoritmo de sumarização utilizado, no caso “RSumm System” (de Ribaldo et al., 2012, introduzido na revisão literária), e botões que disparam as seguintes ações: realizar nova sumarização, ir para a tela inicial do sistema, imprimir o sumário e ajustar configurações extras, como tamanho da fonte. Na área de visualização do sumário, localizada à esquerda no sistema, encontra-se o sumário produzido pelo sistema *RSumm* e informações extras sobre o seu conteúdo, que são: imagens ilustrativas recuperadas de algum repositório de imagens online e informações sobre o conteúdo do sumário, como o número de palavras e os seus termos mais relevantes. Na área de visualização dos textos fonte, localizada à direita no sistema, o

usuário pode visualizar todos os documentos que serviram de entrada para o algoritmo de sumarização, os quais são dispostos em abas que se expandem para a visualização do conteúdo textual. Além do texto fonte, informações extras sobre estes também podem ser visualizadas, como a data e hora de publicação e o nome da publicação, por exemplo, “Globo.com”.

Este sistema foi projetado para que o usuário possa ter uma visualização concorrente do sumário e seus textos fonte na mesma janela de visualização. Essa característica facilita e agiliza o processo de aquisição de conhecimento, pois permite que o usuário alterne o foco de leitura entre o sumário e os textos fonte, rapidamente. A seguir, apresenta-se a avaliação de usabilidade deste sistema com usuários reais.

3.4. Avaliação de Usabilidade do Sistema ViSUM

O projeto do sistema ViSUM envolveu, desde o início, estudos sobre os comportamentos adotados por usuários de sistemas de SAM. Com o objetivo de concluir este estudo, foi realizada uma avaliação da usabilidade do sistema, com o objetivo de avaliar as funcionalidades do sistema, identificar possíveis problemas específicos em sua execução e avaliar o efeito da interface junto aos usuários. Para atender estes objetivos existem diversos métodos de avaliação, que podem ser divididos em dois grupos (Rocha e Baranauskas, 2000): métodos de inspeção de usabilidade, os quais não envolvem usuários e métodos de teste de usabilidade, que incluem métodos experimentais ou empíricos, métodos observacionais e técnicas de questionamento.

Os principais métodos de inspeção de usabilidade existentes são: avaliação heurística, onde é feita a inspeção da interface tendo como base uma pequena lista de heurísticas de usabilidade; revisão de *guidelines*, onde a interface é analisada no sentido de verificar se está de acordo com uma lista de *guidelines* de usabilidade, que geralmente são bem difundidas e utilizadas; inspeção por consistência, em que o avaliador verifica a consistência dentro de uma família de interfaces, quanto à terminologia, cores, *layout*, e tudo mais dentro da interface; e, por fim, percurso cognitivo, em que o avaliador simula a

utilização do sistema por algum usuário típico, “caminhando” na interface para executar as suas tarefas.

Entretanto, neste projeto, optamos por realizar uma avaliação envolvendo usuários reais do sistema, pois, além da avaliação de sua interface, é importante a verificação do impacto das funcionalidades na tarefa de leitura dos sumários multidocumento. Neste caso, temos disponíveis dois métodos de teste de usabilidade com usuários: o método “pensando em voz alta”, que solicita aos usuários que verbalizem tudo o que pensam enquanto utilizam o sistema, de forma que seus pensamentos mostrem como o usuário interpreta cada item da interface; e o método de medidas de performance, que busca a coleta de dados quantitativos sobre a experiência dos usuários com o sistema. Para a avaliação do sistema ViSUM, foi utilizado o método de medidas de performance, por ter uma abrangência maior, pois pode ser aplicado a um número maior de usuários.

O teste foi aplicado com dois tipos diferentes de usuários. O primeiro grupo, denominado “Usuários Comuns”, caracteriza-se por conter pessoas que não têm contato com a área de pesquisa de sumarização automática. São indivíduos que desejam utilizar o sistema unicamente para obter informações de notícias do cotidiano. O segundo grupo, denominado “Usuários Especialistas”, caracteriza-se por conter pesquisadores da área de pesquisa de sumarização automática, os quais já estão acostumados com sistemas da área de PLN. O seu foco é a utilização do sistema em um nível mais profundo, analisando com mais detalhes as funcionalidades, e o seu ganho de informação. Possivelmente, podem levar em consideração, a relevância deste sistema para os seus próprios trabalhos de sumarização.

O grupo dos “usuários comuns” contém 5 indivíduos, de ambos os sexos, que estão em uma faixa etária que corresponde dos 20 aos 30 anos. O grupo dos “usuários especialistas” também contém 5 indivíduos, de ambos os sexos, que estão em uma faixa etária que corresponde dos 20 aos 30 anos. Os testes foram divididos em três partes: primeiramente, é solicitado aos usuários que visualizem dois conjuntos de textos e seus respectivos sumários, previamente estabelecidos para o teste, sem a utilização de nenhum sistema de visualização específico, por meio da visualização destes em arquivos de texto distintos, abertos em janelas do sistema operacional; na segunda parte, os usuários realizam

o upload dos mesmos conjuntos de textos e sumários no sistema ViSUM, e fazem testes livres de todas as funcionalidades e interações possíveis do sistema; por fim, os usuários executam uma busca por algum termo específico, sobre o qual deseja obter um sumário sobre as notícias relacionadas. Após estes experimentos, os usuários preenchem um questionário com perguntas sobre a importância de cada funcionalidade do sistema, as suas interações e o *layout* de sua interface. As únicas informações passadas para os usuários neste teste são as funcionalidades presentes na tela inicial do sistema ViSUM, onde pode ser encontrado a opção de ajuda do sistema, que contém as informações necessárias para a sua utilização. Este questionário está disponível no Apêndice A desta monografia.

A seguir, na Tabela 4 encontram-se os resultados obtidos no preenchimento das perguntas sobre a importância das funcionalidades obtidas no sistema ViSUM (item 1 do questionário). As funcionalidades puderam ser classificadas como: (1) essenciais para o sistema; (2) muito importantes; (3) importantes; (4) pouco importantes; ou (5) não importantes. A classificação 6 se refere à quantidade de pessoas que não souberam avaliar aquela funcionalidade. Na Tabela 4, temos três resultados para cada possível avaliação de cada funcionalidade. O primeiro é referente às avaliações dadas pelos usuários comuns do sistema, o segundo se refere às avaliações dos usuários especialistas e, por fim, o terceiro resultado é referente à soma das avaliações de todos os usuários. Todos os resultados estão dispostos em forma de porcentagem do número de usuários de cada grupo que classificaram cada funcionalidade. Por exemplo, a funcionalidade “i. Visualização do conteúdo de todos os textos sumarizados” teve 60% dos avaliadores comuns a classificando como “1. Essencial para o sistema” e 80% dos avaliadores especialistas também a classificando como “1. Essencial para o sistema”, dando um total de 70% de todos os avaliadores a classificando deste modo.

Tabela 4: Avaliação da importância das funcionalidades do sistema ViSUM

Funcionalidade/Avaliação	1	2	3	4	5	6
i. Visualização do conteúdo de todos os textos sumarizados	60%	20%	20%	-	-	-
	80%	-	20%	-	-	-
	70%	10%	20%	-	-	-
ii. Referências aos textos fonte, a partir das sentenças do sumário	60%	40%	-	-	-	-
	60%	40%	-	-	-	-
	60%	40%	-	-	-	-
iii. Parametrização do tamanho do sumário	20%	20%	20%	20%	-	20%
	-	60%	20%	20%	-	-
	10%	40%	20%	20%	-	10%
iv. Exibição de imagens ilustrativas do sumário	80%	-	20%	-	-	-
	20%	20%	40%	20%	-	-
	50%	10%	30%	10%	-	-
v. Exibição do número de palavras do sumário	-	20%	20%	60%	-	-
	-	40%	-	40%	20%	-
	-	30%	10%	50%	10%	-
vi. Exibição das palavras chave do sumário	40%	20%	40%	-	-	-
	-	80%	20%	-	-	-
	20%	50%	30%	-	-	-
vii. Possibilidade de impressão do sumário e de seus textos fonte	-	20%	60%	20%	-	-
	-	-	80%	20%	-	-
	-	10%	70%	20%	-	-

Pelos resultados obtidos na Tabela 4, pode-se classificar as funcionalidades contidas no sistema ViSUM com relação a importância delas para um sistema de SAM. A funcionalidade “ii. Referências aos textos fonte, a partir das sentenças do sumário” pode ser considerada a mais importante, pois 100% dos avaliadores a classificaram como “Essencial” ou “Muito Importante” para o sistema. A seguir, a funcionalidade que pode ser considerada como a segunda mais importante é “i. Visualização do conteúdo de todos os textos sumarizados”, que obteve 80% das avaliações como “Essencial” ou “Muito Importante” para o sistema.

O questionário aplicado também solicitou aos avaliadores a sugestão de possíveis funcionalidades, não presentes no sistema ViSUM, que julgassem adequadas para um sistema de SAM. As respostas mais interessantes foram: a exibição da data de publicação dos textos fonte, a possibilidade de exclusão de textos fonte na tarefa de sumarização e o destaque das palavras chave no sumário e nos textos fonte.

Na Tabela 5, a seguir, temos a avaliação da facilidade de uso das funcionalidades que requerem alguma interação com o usuário. A sua facilidade pôde ser classificada como: (1) muito fácil de usar; (2) fácil de usar; (3) um pouco confuso de usar; (4) difícil de usar; ou (5) difícil e confuso de usar. O item 6 se refere a quantidade de pessoas que não souberam avaliar a facilidade de uso daquela interação. Do mesmo modo como foram divididos os resultados na Tabela 4, a Tabela 5 a seguir nos apresenta os resultados dos dois grupos de avaliadores, comuns e especialistas, e os resultados da soma das avaliações para cada interação, respectivamente.

Tabela 5: Avaliação da facilidade de uso das interações do sistema ViSUM

Interação/Avaliação	1	2	3	4	5	6
i. Visualização das sentenças em seu texto fonte, quando são clicadas no sumário	60% 80% 70%	40% 20% 30%	- - -	- - -	- - -	- - -
ii. Visualização de um texto-fonte, quando se pressiona o mouse em seu título	20% 60% 40%	80% 20% 50%	- 20% 10%	- - -	- - -	- - -
iii. Parametrização do tamanho do sumário na tela inicial	40% - 20%	20% 60% 40%	40% 40% 40%	- - -	- - -	- - -
iv. Funcionalidade de impressão do sumário e de seus textos fonte	80% 80% 80%	20% 20% 20%	- - -	- - -	- - -	- - -
v. Funcionalidade de upload de arquivos de textos fonte e sumários já produzidos.	40% 40% 40%	40% 20% 30%	20% 40% 30%	- - -	- - -	- - -

Pelos resultados contidos na Tabela 5, nota-se que a grande maioria das interações implementadas no sistema ViSUM foram bastante aceitas pelos avaliadores. Apenas as interações de números iii e v apresentaram problemas para uma parte dos usuários.

Além das questões contidas na Tabela 5, outras duas perguntas referentes à interatividade do sistema ViSUM foram feitas aos avaliadores. Primeiramente, perguntou-se a respeito do ganho de informação na leitura do sumário, quando são utilizadas as referencias aos textos fonte (por meio do clique na sentença do sumário). 90% dos avaliadores consideram que o ganho de informação aumenta com esta funcionalidade (50% deles consideram que o ganho aumenta consideravelmente). A segunda pergunta feita foi

em relação com a facilidade de leitura do sumário quando é utilizado o sistema ViSUM, comparando com uma leitura das mesmas informações (sumário e textos fonte) em arquivos de texto separados. 100% dos avaliadores concordam que a leitura é muito facilitada com a utilização do sistema ViSUM.

Algumas perguntas dissertativas sobre a interatividade do sistema ViSUM também foram realizadas, as quais detectaram alguns problemas na interface. Os problemas mais comuns foram: a maneira que é selecionada a taxa de compressão do sumário; algumas falhas na descrição de ajuda, que pode se tornar confusa para usuários comuns; e a visualização das palavras chave, que estão dispostas sem uma formatação adequada. Esses pontos devem ser adequados em versões futuras da interface.

Na Tabela 6, a seguir, temos a avaliação das características de layout do sistema ViSUM. Cada característica pôde ser classificada como (1) excelente; (2) muito boa; (3) boa; (4) razoável; ou (5) ruim. Seguindo o mesmo padrão de apresentação dos resultados das Tabelas 4 e 5, a Tabela 6 nos apresenta os resultados de cada um dos grupos de usuários que avaliaram o sistema (usuários comuns e especialistas) e também a soma do número de avaliações que cada funcionalidade teve para cada classificação possível. Nota-se que o sistema ViSUM não apresentou grandes problemas em seu layout, na avaliação dos usuários.

Tabela 6: Avaliação das características de layout do sistema ViSUM

Característica do layout/Avaliação	1	2	3	4	5
i. Cores do sistema	60%	40%	-	-	-
	100%	-	-	-	-
	80%	20%	-	-	-
ii. Uso de fontes do sistema	100%	-	-	-	-
	100%	-	-	-	-
	100%	-	-	-	-
iii. Disposição do menu do topo	80%	-	20%	-	-
	100%	-	-	-	-
	90%	-	10%	-	-
iv. Disposição do sumário e dos textos fonte	80%	-	20%	-	-
	60%	20%	20%	-	-
	70%	10%	20%	-	-

3.5. Dificuldades, Limitações e Trabalhos Futuros

A maior dificuldade para a realização deste trabalho foi a falta de conhecimento em programação web por parte do aluno. Este conhecimento foi sendo adquirido no decorrer do trabalho, pois o aluno cursou em paralelo ao seu desenvolvimento duas disciplinas voltadas para este assunto: Técnicas de Programação Web e Sistemas Interativos Web. A maior limitação do trabalho foi o tempo, pois como se trata de um sistema web, é interessante a constante atualização deste, a qual provê melhorias contínuas na interatividade do sistema, e, como este trabalho tem um período definido de um semestre letivo, estas melhorias não são possíveis. Como trabalhos futuros, temos a implementação de outras possíveis funcionalidades para o sistema, como as mencionadas pelos avaliadores na Seção 3.4 e também estudadas na revisão literária deste trabalho.

3.6. Considerações Finais

Nesta Seção, foi relatado todo o sistema desenvolvido neste projeto de conclusão de curso, o qual teve como base para a sua implementação o estudo feito sobre as funcionalidades encontradas nos sistemas de SAM contidos na revisão literária. Com isso temos um ciclo fechado de pesquisa, no qual foram investigadas soluções para a resolução de um problema específico. O autor deste trabalho acredita que a correta implementação deste processo de pesquisa faz com que os trabalhos tenham qualidade e possam ter a base necessária para o sucesso das ações que são propostas inicialmente.

CAPÍTULO 4: CONCLUSÃO

4.1. Contribuições

As contribuições deste trabalho podem ser divididas em dois grupos: primeiramente, este trabalho nos deixa um estudo embasado sobre as funcionalidades existentes em sistemas de SAM, com foco nos usuários destes sistemas. Além disso, como produto final deste projeto, foi desenvolvido um sistema de visualização de SAM, o qual contém algumas das principais funcionalidades estudadas no primeiro momento.

Além do estudo contido na revisão literária e do sistema desenvolvido, também foi realizada uma avaliação não somente do o sistema proposto, mas também da importância das funcionalidades presentes neste. Nota-se, pelos resultados da Seção 3.4, que os usuários que testaram o sistema acreditam que o uso de funcionalidades de visualização é de suma importância em sistemas de SAM, pois nos proveem maior facilidade de uso e maior ganho de informação no processo de leitura do sumário.

Para o aluno, este trabalho foi de grande valia em seu desenvolvimento acadêmico e profissional, pois teve contato com o processo de pesquisa e desenvolvimento de software, que lhe será útil em seu futuro profissional, que pode ser tanto no mercado de trabalho quanto na área acadêmica.

4.2. Considerações sobre o Curso de Graduação

Neste trabalho foram aplicados os conhecimentos adquiridos ao longo da graduação pelo aluno. As disciplinas iniciais de Introdução à Ciências de Computação I e II e Algoritmos e Estruturas de Dados I e II serviram como base para o desenvolvimento de da lógica aplicada nos programas desenvolvidos neste trabalho. Outras disciplinas, como Programação Orientada à Objetos, Análise e Projeto Orientado à Objetos e Engenharia de Software I e II também foram úteis nas fases de projeto e implementação do sistema. Pode-se citar três disciplinas como as mais relevantes para o desenvolvimento deste trabalho: Sistemas Interativos Web e Técnicas de Programação Web, nas quais foram apresentados

os conceitos de programação web, as tecnologias disponíveis e as possibilidades de interação com o usuário; e Interação Humano-Computador, que introduziu o aluno na área, e forneceu conhecimento sobre usabilidade em sistemas interativos e técnicas de avaliação de interfaces, as quais foram utilizadas com bastante intensidade neste trabalho. Por fim, também pode-se citar a disciplina de Metodologia de Pesquisa, a qual introduziu o processo de pesquisa científica ao aluno, que pôde aplica-las na realização deste trabalho.

Sobre o curso de graduação, acredita-se que seu caráter generalista contribui para a formação de conceitos que podem ser aplicados em diferentes tecnologias, e esta característica é de suma importância para o sucesso do aluno, pois se torna capaz de resolver problemas em qualquer escopo da área.

REFERÊNCIAS

CASTRO JORGE, M.L.R.; AGOSTINI, V.; PARDO, T.A.S. Multi-document Summarization Using Complex and Rich Features. In *Anais do VIII Encontro Nacional de Inteligência Artificial*, p. 1-12, 2011.

EVERT, S. A lightweight and efficient tool for cleaning Web pages. In the *Proceedings of the 6th International Conference on Language Resources and Evaluation (LREC 2008)*, 2008.

INOUYE, D.; KALITA, J.K. Comparing Twitter Summarization Algorithms for Multiple Post Summaries. In the *Proceedings of SocialCom/PASSAT*, p. 298-306, 2001.

LEUSKI, A.; LIN, C.Y.; HOVY, E.H. iNeATS: Interactive Multi-Document Summarization. In the *Proceedings of ACL*, p. 125-128, 2003.

MANI, I. *Automatic Summarization*. John Benjamins Publishing Co., Amsterdam, 2001.

MANI, I.; MAYBURY, M.T. *Advances in automatic text summarization*. MIT Press, Cambridge, MA, 1999.

MCKEOWN, K.R.; BARZILAY, R.; EVANS, D.K.; HATZIVASSILOGLOU, V.; KLAVANS, J.L.; NENKOVA, A.; SABLE, C.; SCHIFFMAN, B.; SIGELMAN, S. Tracking and summarizing news on a daily basis with columbia's newsblaster. In the *Proceedings of the Human Language Technology Conference*, p. 280-285, 2002.

MCKEOWN, K.; RADEV, D.R. Generating summaries of multiple news articles. In the *Proceedings of the 18th Annual International ACM-SIGIR Conference on Research and Development in Information Retrieval*, p. 74-82, 1995.

PAULOVICH, F.V.; OLIVEIRA, M.C.F.; MINGHIM, R. The Projection Explorer: A flexible tool for projection-based multidimensional visualization. In the *Proceedings of XX Brazilian Symposium on Computer Graphics and Image Processing (SIBGRAPI)*, p. 27-36, 2007.

RADEV, D., JING, H., STYS, M., TAM, D. Centroid-based summarization of multiple documents. In the *Proceedings Information Processing and Management*, p. 919-938, 2004.

RADEV, D.R. and MCKEOWN, K.R. Generating natural language summaries from multiple on-line sources. *Computational Linguistics*, v. 24, n. 3, p. 469-500, 1998.

RIBALDO, R.; AKABANE, A.T.; PARDO, T.A.S. Multi-document Summarization with Graph Metrics. In the (on-line) *Proceedings of the PROPOR 2012 Demonstrations*, p. 1-3, 2012.

ROCHA, H.V. da; BARANAUSKAS, M.C.C. *Design e Avaliação de Interfaces Humano-Computador*. São Paulo: IME-USP, 2000.

SALTON, G.; MCGILL, M.J. *Introduction to modern information retrieval*. McGraw-Hill, 1986.

SHARIFI, B.; HUTTON, M.A.; KALITA, J.K. Experiments in Microblog Summarization. In the *IEEE Proceedings of 'SocialCom/PASSAT'*, p. 49-56, 2010.

WANG, D.; LI, T.; ZHU, S. SumView: A Web-Based Engine for Summarizing Product Reviews and Customer Opinions. *Expert Systems and Applications*, v. 40, p. 27-33, 2013.

APÊNDICE A – Questionário de Avaliação do Sistema ViSUM

Avaliação de Usabilidade – ViSUM

1. Funcionalidades do sistema ViSUM
 - a. Classifique a importância das funcionalidades presentes no sistema ViSUM:
 - i. Visualização do conteúdo de todos os textos sumarizados

<input type="checkbox"/> 1. Essencial para o sistema	<input type="checkbox"/> 4. Pouco importante
<input type="checkbox"/> 2. Muito importante	<input type="checkbox"/> 5. Não importante
<input type="checkbox"/> 3. Importante	<input type="checkbox"/> 6. Não sei responder
 - ii. Referências dos textos fonte, a partir das sentenças do sumário

<input type="checkbox"/> 1. Essencial para o sistema	<input type="checkbox"/> 4. Pouco importante
<input type="checkbox"/> 2. Muito importante	<input type="checkbox"/> 5. Não importante
<input type="checkbox"/> 3. Importante	<input type="checkbox"/> 6. Não sei responder
 - iii. Parametrização do tamanho do sumário

<input type="checkbox"/> 1. Essencial para o sistema	<input type="checkbox"/> 4. Pouco importante
<input type="checkbox"/> 2. Muito importante	<input type="checkbox"/> 5. Não importante
<input type="checkbox"/> 3. Importante	<input type="checkbox"/> 6. Não sei responder
 - iv. Exibição de imagens ilustrativas do sumário

<input type="checkbox"/> 1. Essencial para o sistema	<input type="checkbox"/> 4. Pouco importante
<input type="checkbox"/> 2. Muito importante	<input type="checkbox"/> 5. Não importante
<input type="checkbox"/> 3. Importante	<input type="checkbox"/> 6. Não sei responder
 - v. Exibição do número de palavras do sumário

<input type="checkbox"/> 1. Essencial para o sistema	<input type="checkbox"/> 4. Pouco importante
<input type="checkbox"/> 2. Muito importante	<input type="checkbox"/> 5. Não importante
<input type="checkbox"/> 3. Importante	<input type="checkbox"/> 6. Não sei responder
 - vi. Exibição das palavras chave do sumário

<input type="checkbox"/> 1. Essencial para o sistema	<input type="checkbox"/> 4. Pouco importante
<input type="checkbox"/> 2. Muito importante	<input type="checkbox"/> 5. Não importante
<input type="checkbox"/> 3. Importante	<input type="checkbox"/> 6. Não sei responder
 - vii. Possibilidade de impressão do sumário e de seus textos fonte

<input type="checkbox"/> 1. Essencial para o sistema	<input type="checkbox"/> 4. Pouco importante
<input type="checkbox"/> 2. Muito importante	<input type="checkbox"/> 5. Não importante
<input type="checkbox"/> 3. Importante	<input type="checkbox"/> 6. Não sei responder

- b. Você gostaria que o sistema ViSUM apresentasse alguma outra funcionalidade? Quais?

2. Interatividade

- a. Classifique a facilidade de uso das interações a seguir

- i. Visualização das sentenças em seu texto fonte, quando são clicadas no sumário

- 1. Muito fácil de usar
- 2. Fácil de usar
- 3. Um pouco confuso de usar
- 4. Difícil de usar
- 5. Muito difícil e confuso de usar
- 6. Não sei responder

- ii. Visualização de um texto-fonte, quando se pressiona o mouse em seu título

- 1. Muito fácil de usar
- 2. Fácil de usar
- 3. Um pouco confuso de usar
- 4. Difícil de usar
- 5. Muito difícil e confuso de usar
- 6. Não sei responder

- iii. Parametrização do tamanho do sumário na tela inicial

- 1. Muito fácil de usar
- 2. Fácil de usar
- 3. Um pouco confuso de usar
- 4. Difícil de usar
- 5. Muito difícil e confuso de usar
- 6. Não sei responder

- iv. Funcionalidade de impressão do sumário e de seus textos fonte

- 1. Muito fácil de usar
- 2. Fácil de usar
- 3. Um pouco confuso de usar
- 4. Difícil de usar
- 5. Muito difícil e confuso de usar
- 6. Não sei responder

- v. Funcionalidade de upload de arquivos de textos fonte e sumários já produzidos.

- 1. Muito fácil de usar
- 2. Fácil de usar
- 3. Um pouco confuso de usar
- 4. Difícil de usar
- 5. Muito difícil e confuso de usar
- 6. Não sei responder

- b. Classifique o ganho de informação na leitura do sumário, quando são utilizadas as referências aos textos fonte
- 1. Aumenta consideravelmente o ganho de informação
 - 2. Aumenta o ganho de informação
 - 3. Aumenta somente um pouco o ganho de informação
 - 4. Não interfere no ganho de informação
 - 5. Atrapalha o ganho de informação
 - 6. Não sei responder

- c. Classifique o quanto facilita a leitura do sumário quando é utilizado o sistema ViSUM, comparando com uma leitura das informações em arquivos de texto distintos.
- 1. Facilita muito a leitura
 - 2. Facilita a leitura
 - 3. Não há melhoras
 - 4. Dificulta a leitura
 - 5. Dificulta muito a leitura
 - 6. Não sei responder

- d. Comente sobre a experiência de upload de arquivos de textos fonte e sumários já produzidos. Há alguma coisa que você gostaria que fosse diferente?

- e. Comente sobre a experiência de utilização das funcionalidades do sistema ViSUM. Há alguma coisa que você gostaria que fosse diferente?

3. Características de layout do sistema ViSUM

a. Classifique o design das telas do sistema ViSUM

i. Cores do sistema

- Excelente
- Muito boa
- Boa
- Razoável
- Ruim

ii. Uso de fontes do sistema

- Excelente
- Muito boa
- Boa
- Razoável
- Ruim

iii. Disposição do menu do topo

- Excelente
- Muito boa
- Boa
- Razoável
- Ruim

iv. Disposição do sumário e dos textos fonte

- Excelente
- Muito boa
- Boa
- Razoável
- Ruim

b. Comente sobre o layout do sistema. Quais são seus pontos fortes e fracos?
