

**UNIVERSIDADE FEDERAL DO ESPÍRITO SANTO  
CENTRO TECNOLÓGICO  
PROGRAMA DE PÓS-GRADUAÇÃO EM INFORMÁTICA**

**FELICIA DE LUCIA CASTILLO**

**TÉCNICAS PARA VISUALIZAR INFORMAÇÃO E EXTRAIR CONHECIMENTO DE  
BASES DE INFORMAÇÕES: ESTUDO DE CASO DAS PRODUÇÕES DOS  
PROFESSORES DO PPGI-UFES**

**VITÓRIA  
2017**

**FELICIA DE LUCIA CASTILLO**

**TÉCNICAS PARA VISUALIZAR INFORMAÇÃO E EXTRAIR CONHECIMENTO DE  
BASES DE INFORMAÇÕES: ESTUDO DE CASO DAS PRODUÇÕES DOS  
PROFESSORES DO PPGI-UFES**

Dissertação apresentada ao Programa de Pós-Graduação em Informática do Centro Tecnológico da Universidade Federal do Espírito Santo, como requisito parcial para obtenção do Grau de Mestre em Informática.

Orientador: Prof. Dr. Celso Alberto Saibel Santos.

**VITÓRIA  
2017**

## FICHA CATALOGRÁFICA

De Lucia Castillo, Felicia.

Técnicas para Visualizar Informação e Extrair Conhecimento de Bases de Informações: Estudo de Caso da Produção dos Professores do PPGI-UFES – Vitória, 2017.

Nº de páginas:

Área de concentração: Redes de Computadores e Sistemas Distribuídos.

Orientador: Prof. Celso Alberto Saibel Santos.

Dissertação (Mestrado em Informática) – Universidade Federal do Espírito Santo, Centro Tecnológico.

1. Visualização; 2. Informação; 3. Técnicas

**FELICIA DE LUCIA CASTILLO**

**TÉCNICAS PARA VISUALIZAR INFORMAÇÃO E EXTRAIR CONHECIMENTO DE BASES DE INFORMAÇÕES: ESTUDO DE CASO DAS PRODUÇÕES DOS PROFESSORES DO PPGI-UFES**

Dissertação submetida ao programa de Pós-Graduação em Informática do Centro Tecnológico da Universidade Federal do Espírito Santo, como requisito parcial para a obtenção do Grau de Mestre em Informática.

Aprovada em 27 de abril de 2017.

**COMISSÃO EXAMINADORA**

---

Prof. Dr. Celso Alberto Saibel Santos  
Universidade Federal do Espírito Santo  
PPGI - Orientador

---

Prof. Dr. Patrick Marques Ciarelli  
Universidade Federal do Espírito Santo  
PPGEE - Avaliador Externo

---

Prof. Dr. Vítor Estêvão Silva Souza  
Universidade Federal do Espírito Santo  
PPGEE - Avaliador Externo

## **DEDICATÓRIA**

Dedico o trabalho a todos os pesquisadores que trabalham na área de Visualização de Informação.

## **AGRADECIMENTOS**

Gostaria de agradecer à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), à Universidade Federal do Espírito Santo (UFES) e à Organización de Estados Americanos (OEA) pelo apoio financeiro.

**“Separación:** La lógica es pensamiento. Los efectos son acción. Por lo tanto, el sabio piensa antes de actuar, e actúa solo cuando el pensamiento está hecho. Si intentas realizar efectos e lógica, al mismo tiempo, puedes crear efectos secundarios ocultos que causen errores en la lógica. Mantén funciones pequeñas. Haz una cosa a la vez, e hazla bien”

James Sinclair.

## SUMÁRIO

LISTA DE ILUSTRAÇÕES E TABELAS.....	IX
RESUMO.....	XI
ABSTRACT .....	XII
1 CONTEXTO E MOTIVAÇÃO.....	14
1.1 VISUALIZAÇÃO DE INFORMAÇÃO .....	15
1.2 PROBLEMAS RELACIONADOS.....	16
1.3 OBJETIVOS E CONTRIBUIÇÕES .....	17
1.4 PROCESSO DE DESENVOLVIMENTO .....	18
1.5 ORGANIZAÇÃO DO TRABALHO .....	18
2 EXTRAÇÃO DE CONHECIMENTO DAS INFORMAÇÕES .....	21
2.1 PROCESSOS DE VISUALIZAÇÃO DE INFORMAÇÃO.....	22
2.2 MINERAÇÃO DE INFORMAÇÃO E DESCOBERTA DE CONHECIMENTO ...	25
2.2.1 Processo de Descoberta de Conhecimento .....	25
2.3 TÉCNICAS DE VISUALIZAÇÃO DE INFORMAÇÃO .....	26
2.3.1 Visualização de Informação em Nuvem de Palavras .....	28
2.3.2 Visualização em Nuvem de Palavras Animadas .....	29
2.3.3 Visualização em Grafos.....	30
2.3.4 Visualização em Gráfico de Barras .....	31
2.3.5 Visualização em Gráfico Circular .....	32
2.4 CONSIDERAÇÕES DO CAPÍTULO.....	32
3 TRABALHOS RELACIONADOS .....	35
3.1 MANY EYES .....	35
3.2 ZINGCHART .....	37
3.3 WORDSWARM .....	37
3.4 PYGAME .....	38
3.5 UTILIZANDO TÍTULOS DE ARTIGOS CIENTÍFICOS NA CONSTRUÇÃO DE REDES SEMÂNTICAS PARA CARACTERIZAR ÁREAS DE PESQUISA .	38
3.6 ANÁLISE DA REDE DE COLABORAÇÃO DO SBRC: AS PRIMEIRAS 30 EDIÇÕES .....	39
3.7 VISUALIZANDO 15 ANOS DO IHC .....	40
3.8 WEBMEDIA, 18 ANOS: QUEM SOMOS E O QUE TEMOS FEITO NAS ÚLTIMAS DUAS DÉCADAS?.....	40
3.9 CONSIDERAÇÕES DO CAPÍTULO.....	40



4	A PROPOSTA.....	43
4.1	CENÁRIO REAL DA APLICAÇÃO .....	43
4.2	O “VI-PROJECT” .....	44
4.3	METODOLOGIA DE ANÁLISE E DESENHO DO SISTEMA.....	45
4.3.1	Identificar problemas, oportunidades e objetivos .....	46
4.3.2	Determinação dos requisitos .....	46
4.3.3	Análise das necessidades .....	47
4.3.4	Desenho do sistema.....	47
4.3.5	Desenvolvimento do sistema.....	47
4.3.6	Teste do sistema .....	47
4.3.7	Implementação do sistema.....	47
4.4	CAMADAS DO SISTEMA WEB IMPLEMENTADO .....	47
4.5	DETALHES SOBRE O DESENHO DO SISTEMA .....	48
4.5.1	Processo de Migração da Informação .....	49
4.5.2	Processo de Geração das Visualizações .....	51
4.5.2.1	Visualizações Estatísticas .....	51
4.5.2.2	Nuvem de Palavras Animadas .....	53
4.6	ARQUITETURA DO SISTEMA.....	56
4.6.1	Módulo de Administração dos Usuários .....	57
4.6.2	Módulo dos Professores.....	58
4.6.3	Módulo dos Grupos de Pesquisa .....	59
4.6.4	Módulo das Nuvens de Palavras .....	60
4.6.5	Módulo dos termos da ACM .....	62
5	EXPERIMENTOS E RESULTADOS .....	64
5.1	RESULTADOS DAS VISUALIZAÇÕES .....	64
5.1.1	Nuvens de Palavras .....	65
5.1.2	Rede (Grafos) das Publicações.....	71
5.1.3	Visualizações Estatísticas .....	74
5.1.4	Visualização dos Termos da ACM .....	84
5.2	CONSIDERAÇÕES DO CAPÍTULO.....	86
6	CONCLUSÃO E TRABALHOS FUTUROS .....	88
6.1	PUBLICAÇÕES RELACIONADAS.....	91
	REFERÊNCIAS.....	92
	APENDICE 1 - PLANO DE IMPLANTAÇÃO .....	98

APENDICE 2 - CONCEITOS E FERRAMENTAS .....	99
FERRAMENTAS .....	99
TECNOLOGIAS E LINGUAGENS.....	100
jQuery .....	100
Bootstrap.....	101
HTML .....	101
JavaScript.....	101
CSS .....	102
Apache .....	102
Symfony .....	102
Silex .....	103
PHP .....	103
MySQL .....	104
Python .....	104
Padrão Modelo Visão Controlador (MVC) .....	105
Transferência de Estado Representacional (REST).....	105
Linguagem de Modelagem Unificada (UML) .....	106
FosUserBundle.....	107

## LISTA DE ILUSTRAÇÕES E TABELAS

FIGURA 1 - EXEMPLO DE VISUALIZAÇÃO DE VALORES POR MEIO DE BARRAS .....	21
FIGURA 2 - DIAGRAMA ESQUEMÁTICO DO PROCESSO DE VISUALIZAÇÃO .....	22
FIGURA 3 - PROCESSO DE VISUALIZAÇÃO DE INFORMAÇÃO.....	23
FIGURA 4 - PROCESSO DO KDD.....	26
FIGURA 5 - UMA DAS INTERFACES DO MANY EYES.....	36
FIGURA 6 - CAMADAS DE UM SISTEMA WEB.....	48
FIGURA 7 - PROCESSO DE OBTENÇÃO DOS DADOS A PARTIR DA PLATAFORMA LATTES.....	49
FIGURA 8 - ESQUEMA DO BANCO DE DADOS .....	50
FIGURA 9 - PROCESSO DE CRIAÇÃO DAS VISUALIZAÇÕES ESTATÍSTICAS.....	53
FIGURA 10 - PROCESSO DE CRIAÇÃO DAS NUVENS DE PALAVRAS ANIMADAS.....	55
FIGURA 11 - DIAGRAMA DE PACOTES VI-PROJECT.....	57
FIGURA 12 - DIAGRAMA DE PACOTES MÓDULO ADMINISTRAÇÃO DOS USUÁRIOS.....	58
FIGURA 13 - DIAGRAMA DE PACOTES MÓDULO PROFESSORES.....	59
FIGURA 14 - DIAGRAMA DE PACOTES MÓDULO GRUPO DE PESQUISA.....	60
FIGURA 15 - DIAGRAMA DE PACOTES MÓDULO NUVENS DE PALAVRAS.....	61
FIGURA 16 - DADOS PARA A CRIAÇÃO DAS TÉCNICAS DE VISUALIZAÇÃO DE INFORMAÇÃO .....	65
FIGURA 17 - NUVENS DE PALAVRAS NO <i>SITE</i> VI-PROJECT.....	67
FIGURA 18 - TEMAS PESQUISADOS NO ANO 1998.....	68
FIGURA 19 - TEMAS PESQUISADOS NO ANO 1999.....	68
FIGURA 20 - TEMAS PESQUISADOS NO ANO 2002.....	69
FIGURA 21 - TEMAS PESQUISADOS NO ANO 2006.....	70
FIGURA 22 - TEMAS PESQUISADOS NO ANO 2010.....	70
FIGURA 23 - TEMAS PESQUISADOS NO ANO 2015.....	71
FIGURA 24 - REDE DAS PUBLICAÇÕES DE UM GRUPO DE PESQUISA.....	72
FIGURA 25 - VISUALIZAÇÃO ESTATÍSTICA DE UM GRUPO DE PESQUISA ....	75
FIGURA 26 - VISUALIZAÇÃO PRODUÇÕES BIBLIOGRÁFICAS / TRABALHOS ORIENTADOS 2013.....	76

FIGURA 27 - VISUALIZAÇÃO PRODUÇÕES BIBLIOGRÁFICAS / TRABALHOS ORIENTADOS 2010.....	77
FIGURA 28 - VISUALIZAÇÃO PRODUÇÕES BIBLIOGRÁFICAS / TRABALHOS ORIENTADOS 2011.....	79
FIGURA 29 - VISUALIZAÇÃO PRODUÇÕES BIBLIOGRÁFICAS / TRABALHOS ORIENTADOS 2012.....	80
FIGURA 30 - VISUALIZAÇÃO INFORMAÇÃO PRODUÇÕES BIBLIOGRÁFICAS / TRABALHOS ORIENTADOS.....	81
FIGURA 31 - VISUALIZAÇÃO KEYWORDS DOS PROFESSORES.....	82
FIGURA 32 - VISUALIZAÇÃO KEYWORDS DOS PROFESSORES.....	83
FIGURA 33 - VISUALIZAÇÃO KEYWORDS DOS PROFESSORES.....	84
FIGURA 34 - VISUALIZAÇÃO TERMOS DA ACM.....	85
FIGURA 35 - VISUALIZAÇÃO TERMOS DA ACM.....	86
TABELA 1 - FERRAMENTAS UTILIZADAS NO PROCESSO DE DESENVOLVIMENTO DO SISTEMA.....	99

## RESUMO

A Visualização é uma parte muito importante para conseguir a compreensão real da informação e para extrair conhecimentos relevantes. Portanto, a técnica para visualizar a informação selecionada deve ser compatível com o conteúdo que se vai apresentar. A proposta desta dissertação é apresentar um sistema Web que integre diversas técnicas de visualização criadas, que sejam fáceis, rápidas, entendíveis, dinâmicas e adaptáveis, dos temas desenvolvidos pelos integrantes de um grupo de pesquisa, durante alguns períodos. Essas técnicas são construídas tomando por base principalmente as palavras-chave e títulos das produções bibliográficas, representando assim a evolução dos temas pesquisados ao longo do tempo, entre outras coisas.

**Palavras-chave:** Visualização de Informação, Técnicas de Visualização, Recursos de Visualização, Produções Bibliográficas, Dados, Conhecimento.

## ABSTRACT

Visualization is a very important part of achieving real understanding of information and extracting relevant knowledge. Therefore, the technique to visualize the selected information must be compatible with the content that will be presented. The purpose of this dissertation is to present a Web system that integrates several visualization techniques, which are easy, fast, understandable, dynamic and adaptable, from the topics developed by the members of a research group in a period of time. These techniques are based mainly on the keywords and titles of the bibliographic publications, thus representing the evolution of the topics researched over time, among other things.

**Keywords:** Information Visualization, View Techniques, Preview Features, Bibliographic Publications, Data, Knowledge.

# **1 INTRODUÇÃO**

## 1 CONTEXTO E MOTIVAÇÃO

A quantidade de informação digital produzida tem se tornado cada vez maior; por isso, as técnicas utilizadas para visualizar esta informação de maneira eficiente precisam evoluir constantemente para acompanhar este cenário.

As ferramentas e técnicas de visualização têm o objetivo de tornar mais fácil e rápido o entendimento dos dados apresentados, assim como a qualidade das informações representadas por eles. Elas têm sido criadas para esta finalidade, todavia a demanda por formas adequadas de visualização ainda persiste em diversos cenários. Frequentemente, estes cenários requerem formas de visualizar, de forma simultânea, conjuntos estruturados de informação que foram concebidos para serem visualizados de forma isolada.

A visualização das produções científicas de grupos de pesquisadores, tendo como ponto de partida as informações oriundas da base de CV Lattes do CNPq representa um cenário como o anterior. A base de CV Lattes oferece dados públicos através da *Web*, os quais podem processados, manipulados e analisados a fim de revelar informações sobre a produção científica de pesquisadores. Trata-se de um conjunto bem estruturado de informações textuais, que são organizadas e agrupadas de forma a gerar uma boa visualização por tipo de atividade realizada, porém diversos aspectos poderiam ser observados mais facilmente se houvesse o uso de uma visualização adequada. Não se pode, por exemplo, verificar o crescimento ou a diminuição da produção acadêmica de um pesquisador ao longo do tempo, nem as associações dos seus temas atuais e passados e suas publicações. Também não é possível confrontar informações como a quantidade de orientações concluídas e as publicações associadas a estas orientações e tampouco visualizar de forma coerente informações de múltiplos pesquisadores pertencentes a uma mesma linha de pesquisa.

Desta forma, é importante que se estude como as técnicas atuais de visualização de informação podem ser expandidas e agregadas a fim de encontrar formas alternativas e aprimoradas de exibição, como a geração de visualizações



animadas, incorporando uma dimensão temporal em formas de visualização que atualmente são estáticas, entre outras.

## 1.1 Visualização de Informação

A visualização de informação pode ser definida como o uso de representações visuais interativas de dados abstratos, não físicos, para expandir o conhecimento (CARD; MACKINLAY; SHNEIDERMAN, 1999). A visualização de informação está centrada na extração de estruturas de grandes conjuntos de dados que podem ser representados visualmente aos usuários. Por conseguinte, a visualização de informações envolve diferentes processos. Há uma transformação dos dados brutos em abstrações analíticas que, em seguida, se transformam em um modelo espacial-visual, de modo que através de processos de desenho visual, ou modelo visual é apresentado ao usuário graficamente, visível e compreensivelmente (OLMEDA, 2014).

A visualização de informação cria formas de representações visuais interativas que aperfeiçoam a carga cognitiva na representação visual das estruturas, relações e padrões da informação exibida, assim como permite obter visões analíticas diferentes dos dados contidos nas visualizações. Ela surge em resposta à procura de pesquisadores por ferramentas que favoreçam a análise e compreensão dos dados abstratos através da utilização da computação gráfica interativa e o uso de técnicas de visualização (MARTIG; CASTRO; DI LUCA, 2002) (NUALART; PÉREZ; WHITELAW, 2014).

A visualização de informação tem três objetivos principais: (1) descobrir ou explorar (em busca de relacionamentos, tendências e fenômenos interessantes); (2) apoiar a tomada de decisões (validando ou refutando hipóteses) e (3) explicar (por meio da apresentação de informações para outros) (CARD; MACKINLAY; SHNEIDERMAN, 1999). Ela é caracterizada por transformar dados brutos em informações relevantes, permitindo que os usuários interajam, transformem e interpretem essas informações e adquiram conhecimento sobre os dados a partir dessas ações.

O uso de técnicas de visualização de informação também desperta interesse em questões ligadas ao termo “*big data*”. Neste sentido, a exploração de novas alternativas em estudos de métricas, representação e análise de grandes quantidades de dados, mineração de dados e de texto, recuperação de informação, entre outros, são temas relacionados diretamente à visualização de informação. Ela tem avançado rapidamente em áreas de aplicação como a gestão da informação pessoal ou organizacional, aprendizagem, análise de informações, representação do domínio de conhecimento, etc. (CHEN, 2002). No campo da Tecnologia da Informação, a visualização da informação tem resultados de estudos e aplicação de técnicas emergentes e inovadoras originadas na computação, permitindo novas representações do conhecimento e novos olhares a partir da análise dos dados visualizados (TORRES, 2010).

## **1.2 Problemas Relacionados**

Para muitos usuários, selecionar, processar, analisar e visualizar coleções de informações é uma atividade necessária, mas acaba se tornando árdua devido à quantidade excessiva de informações que devem manipular. Isso exige a criação de novas técnicas para lidar com esse tipo de informação. Em muitos casos, a forma de apresentação é um fator crucial para que o usuário tenha um bom entendimento das informações analisadas (NUALART; PÉREZ; WHITELAW, 2014).

A visualização de informação envolve ainda outras questões, uma vez que se trata de uma abordagem centrada no usuário. Usabilidade, compreensão das principais tarefas perceptivo-cognitivas, conhecimento prévio, métricas de qualidade intrínseca, escalabilidade, estética, inferência visual e previsões, mudanças de paradigmas, visualização de domínios de conhecimento, entre outras são exemplos destes desafios.

Associar métodos e técnicas de visualização para as necessidades e problemas dos usuários com diferentes níveis de habilidades e treinamento; investigar se as ferramentas de visualização são mais eficazes para apoiar a gestão da informação e conhecimento, métodos e critérios de avaliação adequados dos processos realizados; ganhar a compreensão sobre como diferentes grupos de

usuários percebem os resultados alcançados através da aplicação de diferentes técnicas de visualização, qual é a melhor forma de extrair conhecimento das visualizações e ser capaz de propor critérios a considerar em soluções futuras, são outros problemas a serem solucionados na área.

### **1.3 Objetivos e Contribuições**

Em linhas gerais, buscando novas soluções para os problemas citados, esse trabalho tem como objetivo a criação e a avaliação de formas de visualização de informação, geradas a partir de dados textuais, permitindo a detecção de padrões, comportamentos e evidências a partir do dinamismo na apresentação visual dos resultados. O objetivo destas visualizações é ampliar a possibilidade de que fatos escondidos sejam descobertos e melhorar a forma de examinar e compreender essas informações.

As formas de visualização criadas nesta dissertação são disponibilizada por meio de uma plataforma Web aberta, que permite liberdade de execução, redistribuição, estudo ou modificação, incentivando o seu reuso e o desenvolvimento tecnológico na área.

Além do objetivo geral, os seguintes objetivos específicos foram estabelecidos:

- Propor e desenvolver novas formas de visualização de informação para bases de informação textual.
- Descobrir como algumas técnicas de visualização de informação podem trabalhar em conjunto para gerar conhecimento.
- A partir de dados obtidos a partir da plataforma Lattes do CNPq, testar as técnicas de visualização em termos de funcionalidades oferecidas, compreensão do que é apresentado e extração de conhecimento.
- Disponibilizar a solução no ambiente Web, com diversas implementações das formas de visualização escolhidas, aplicadas sobre o conjunto de dados relacionados à produção científica de professores do PPGI da UFES.

## 1.4 Processo de desenvolvimento

O processo de desenvolvimento deste trabalho seguiu as seguintes etapas:

1. No primeiro passo se realizou o levantamento do estado da arte das técnicas e ferramentas de visualização de informação, assim como uma análise sobre o que ainda precisava ser feito.
2. No segundo passo se realizou a identificação de um cenário real de aplicação para modelar um estudo de caso e avaliar o que fosse desenvolvido.
3. O terceiro passo tratou-se da definição do escopo do projeto, o que será desenvolvido e avaliado. Ainda dentro do cenário de aplicação escolhido para o estudo de caso, se definiu quais visualizações e quais ferramentas deveriam ser desenvolvidas.
4. O quarto passo foi o desenvolvimento das técnicas de visualização de informação definidas, assim como a definição das tecnologias a serem utilizadas na implementação, como, por exemplo, linguagens, sistemas, bibliotecas, *frameworks*, etc.
5. O quinto passo foi o teste das implementações num estudo de caso real, juntamente com a análise dos resultados.
6. O sexto passo refere-se à escrita da dissertação e discussão dos resultados alcançados e das limitações do trabalho.

## 1.5 Organização do Trabalho

Após essa introdução, o restante do texto dessa dissertação está organizado em outros cinco capítulos. O Capítulo 2 apresenta a fundamentação teórica necessária para o desenvolvimento do trabalho, abordando aspectos da visualização de informação. O capítulo seguinte descreve os trabalhos relacionados encontrados e uma análise dos mesmos. O Capítulo 4 descreve a arquitetura e implementação da plataforma, a metodologia para o desenvolvimento bem como as ferramentas e tecnologias empregadas no processo. A avaliação da plataforma, assim como os experimentos realizados para a verificação de suas funcionalidades são detalhados no Capítulo 5. O último capítulo apresenta as conclusões e trabalhos futuros que

podem adicionalmente contribuir nas pesquisas no campo de visualização de informação.

ERRO! FONTE DE REFERÊNCIA NÃO ENCONTRADA.  
**VISUALIZAÇÃO DE INFORMAÇÕES**

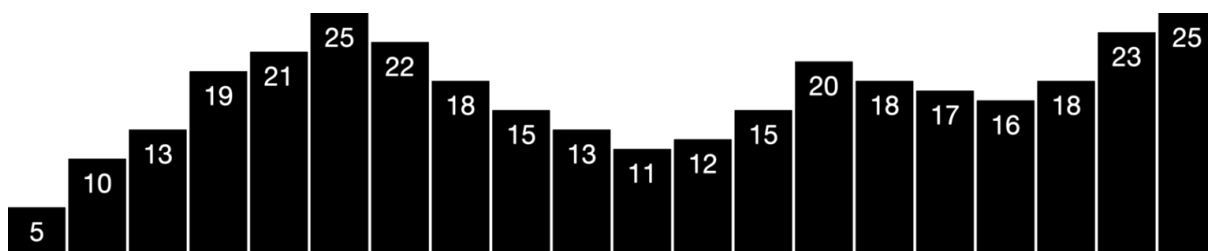
## 2 EXTRAÇÃO DE CONHECIMENTO DAS INFORMAÇÕES

Atualmente, a quantidade de informações manipuladas é excessiva, sobrecarregando as pessoas (MURRAY, 2013). Essas informações se tornam úteis quando são aplicados métodos ou técnicas para conseguir a extração de conhecimento a partir das visualizações.

A visualização de informação pode ser vista como uma representação visual com objetivo de transformar dados simples ou brutos em informação (CARD; MACKINLAY; SHNEIDERMAN, 1999). O seu principal objetivo é transmitir as informações ao usuário de forma clara e eficaz por meio de componentes gráficos. Ela ajuda ao usuário a ver o mundo de outra forma, ajuda a revelar padrões, tendências ou conhecimentos antes não encontrados, simplifica a busca de informações, aumenta os recursos acessíveis dos usuários, entre outras vantagens. Ela é especialmente útil para que seja possível fazer uma análise rápida de uma situação específica. Em outras palavras, a visualização é um processo de mapeamento da informação, com regras que interpretam a informação e expressam seus valores com propriedades visuais (MURRAY, 2013).

No exemplo do gráfico da Figura 1, a forma de visualização utilizada é gerada a partir de uma regra específica: os valores maiores são representados com barras mais altas, sendo facilmente identificados a partir de uma análise qualitativa. Além disso, a representação dos valores na parte superior das barras facilita a análise quantitativa dos dados, permitindo determinar, rapidamente o (s) maior (es) valores do conjunto analisado.

FIGURA 1 - EXEMPLO DE VISUALIZAÇÃO DE VALORES POR MEIO DE BARRAS



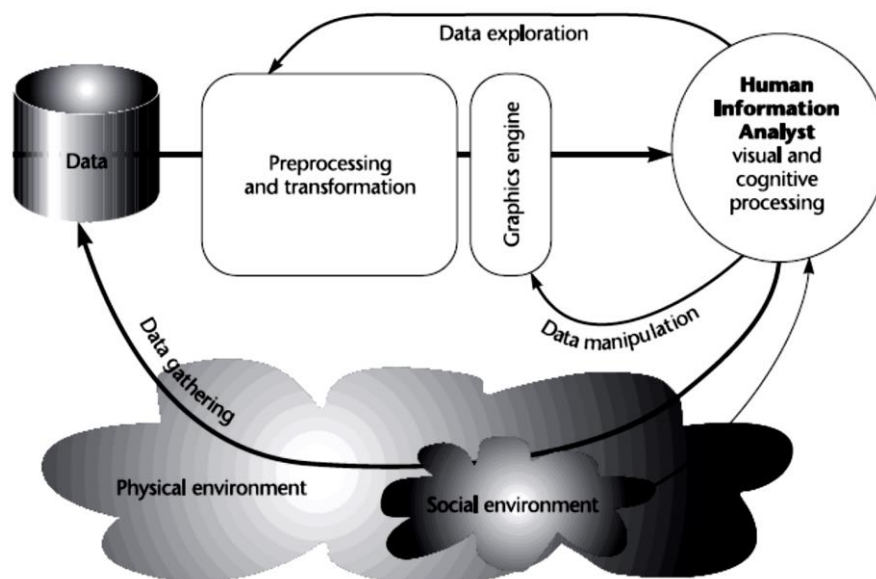
Fonte: Adaptada de (MURRAY, 2013)

## 2.1 Processos de visualização de Informação

De acordo com (WARE, 2012), o processo de visualização de informação inclui quatro etapas básicas, que trabalham em conjunto (Figura 2). As quatro etapas consistem em:

1. Recolher e armazenar as informações.
2. Pré-processar as informações para transformá-las em algo que se pode compreender pela mente humana (conhecimento).
3. Utilizar a plataforma disponível (hardware e software) em conjunto com algoritmos gráficos necessários para produzir a imagem ou visualizações nos dispositivos de apresentação.
4. Utilizar o sistema perceptivo e cognitivo humano (o destinatário) para completar a visualização.

FIGURA 2 - DIAGRAMA ESQUEMÁTICO DO PROCESSO DE VISUALIZAÇÃO



Fonte: Adaptada de (WARE, 2012)

Como pode ser observado na figura anterior, o processo envolve a utilização de um motor gráfico (computador) para a geração das visualizações. Na proposta de Ware (2012), a etapa mais longa e complexa do processo está associada à coleta de informação. Obviamente, sem a informação não faz sentido pensar em visualizações. A coleta de informação está baseada num ambiente físico (fonte de



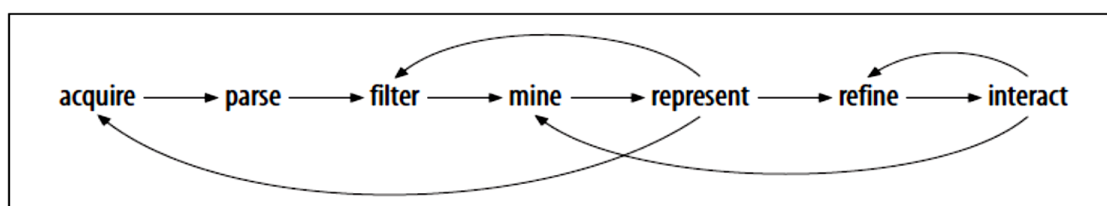
dados disponíveis) e um ambiente social que determina quais informações procurar em função da percepção do usuário. Outra etapa controla o pré-processamento computacional, que é realizado antes da visualização. Esta etapa é importante, pois o significado real da informação poderia ser alterado se houver uma percepção errada do usuário causada por um pré-processamento inadequado sobre os dados. Finalmente, o processo de visualização de informação deve ser altamente interativo, porque é o usuário que determina quando a visualização chega ao fim a partir da interação com o conjunto de dados.

Uma questão-chave no processo de visualização é como determinar a melhor maneira de transformar a informação em algo que as pessoas possam entender para apoiar a tomada de decisão eficiente.

Para Fry (2007) o processo de visualização de informação consiste em uma série de passos que respondem uma pergunta que origina a necessidade de visualização de informação (Figura 3). Os passos consistem em:

- **Adquirir:** Obter a informação (arquivo, disco ou na rede)
- **Analisar:** Prover uma estrutura à informação (categorias)
- **Filtrar:** Apagar as informações que não sejam importantes
- **Extrair:** Aplicar métodos estatísticos de mineração de informação para encontrar padrões
- **Representar:** Selecionar um modelo visual básico
- **Refinar:** Melhorar o modelo visual básico para torná-lo visualmente mais atrativo e útil
- **Interagir:** Agregar métodos para manipular a informação ou determinar a informação que deve ser apresentada

FIGURA 3 - PROCESSO DE VISUALIZAÇÃO DE INFORMAÇÃO



Fonte: Adaptada de (FRY, 2007)

Em (CARD; MACKINLAY; SHNEIDERMAN, 1999) também é proposto um modelo do processo para a Visualização de Informação:

- **Transformação dos dados:** Os dados primários associados a fenômenos complexos da realidade se transformam em uma visão estruturada, através da aplicação de técnicas estatísticas para o processamento de dados, com o objetivo de fazer redução dimensional de um conjunto de dados, computar parâmetros, extrair características da informação, etc. Os dados podem corresponder a fenômenos físicos ou abstratos, espaciais ou temporais, e sua complexidade, contexto e propósito vão determinar o tipo de representação que permita comunicação eficaz.
- **Mapeamento Visual:** É necessária a definição das estruturas visuais apropriadas dos dados transformados. Para isso, são considerados a organização e distribuição espacial dos dados, os objetos gráficos (pontos, linhas, áreas) utilizados e seus atributos (posição, dimensão, cor, forma) que melhor se adaptem à percepção humana e, como representar as conexões entre os diferentes componentes visuais que correspondam com as relações dos dados. A representação das estruturas visuais não deve ter uma interpretação ambígua.
- **Transformação das Visualizações:** As transformações das visualizações modificam interativamente as estruturas visuais ao estabelecer parâmetros com a finalidade de melhorar a percepção dos dados analisados. Algumas dessas transformações correspondem com a afirmação definida em (SHNEIDERMAN, 1996): *Overview first, zoom and filter, then details-on-demand*.

Ao comparar os três processos propostos, podem se observar pontos em comum, por exemplo: a importância de obter ou ter os dados para eles serem transformados, analisados ou processados, e serem convertidos posteriormente em visualizações. Porém, Fry (2007) ilustra as conexões ao longo do processo para representar o trabalho em equipe e a importância de cada passo, e não tratar a visualização de informação como um processo completamente linear. Também, no processo de Ware (2012) é possível visualizar o trabalho em conjunto das etapas e a especial importância no sistema perceptivo e cognitivo do usuário para interpretar a

visualização da informação após ser gerada. O modelo proposto em (CARD; MACKINLAY; SHNEIDERMAN, 1999) estabelece três passos muito específicos e diretos para gerar as visualizações de informação.

## **2.2 Mineração de Informação e Descoberta de Conhecimento**

Para Fayyad, Piatetsky-Shapiro e Smyth (1996) a Mineração de Informação é a aplicação de algoritmos específicos para extrair padrões. Segundo os autores, a Mineração de Informação é apenas um passo do processo muito mais amplo que é a descoberta de conhecimento, conhecido como KDD (*Knowledge Discovery in Databases*). Sendo assim, KDD refere-se ao processo da descoberta de padrões válidos, novos, potencialmente úteis e compreensíveis nas informações.

A visualização de informação ajuda na gestão do conhecimento ao permitir representar, criar e compartilhar o conhecimento (ZHU; CHEN, 2005). A visualização de informação tem foco na criação, transferência e comunicação do conhecimento, assim como melhorar sua transferência e comunicação entre as pessoas. Além disso, a visualização de informação representa o conhecimento usando metáforas, mapas conceituais, diagramas, entre outros, com o fim de tomar decisões ou resolver problemas. Assim, ela adiciona técnicas e ferramentas para fortalecer os processos de gestão do conhecimento.

No caso específico da análise e mineração de textos, as técnicas, tais como as apresentadas em (FELDMAN; SANGER, 2006), tem como objetivo detectar uma informação que esteja incluída num texto, mas que é desconhecida até então, já que não pôde ser efetivamente descrita (HEARST, 2003). Estas técnicas levam à compreensão do texto de uma forma diferente, em uma dinâmica mais interativa, modificando o conteúdo linear original e transformando-o num conjunto de informações não estruturadas associado a uma representação gráfica adequada.

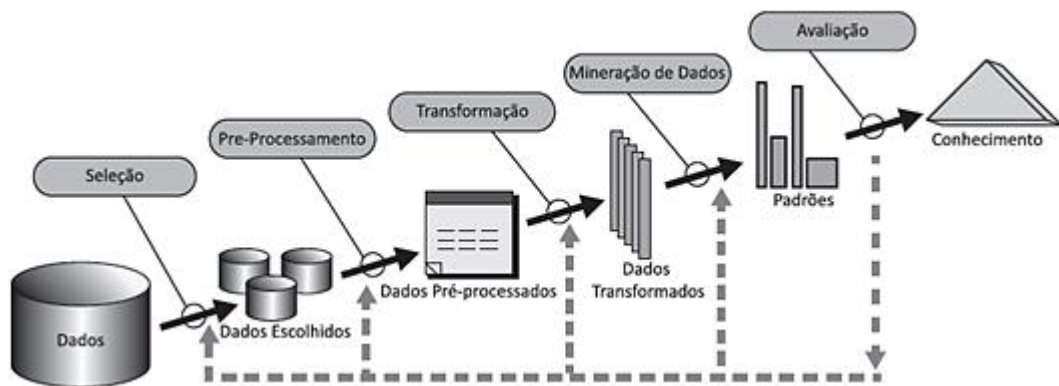
### **2.2.1 Processo de Descoberta de Conhecimento**

Apesar de ter sido proposto há muitos anos, pode-se afirmar que o processo iterativo e iterativo de (FAYYAD; PIATETSKY-SHAPIRO; SMYTH, 1996) ainda

continua válido atualmente. O processo é composto por vários passos, exigindo que decisões sejam tomadas pelos usuários durante a sua execução.

Antes de selecionar o subconjunto de dados a ser analisado, é necessário que se compreenda o domínio de aplicação e o objetivo do processo do KDD. Uma vez feito isso, deve ser selecionada uma parte das informações sobre as quais serão realizadas as tarefas de descoberta de padrões. Estas informações devem ser pré-processadas, com a limpeza dos dados, eliminação de valores extremos, decisões sobre o tratamento dos dados faltantes e dos dados que mudam com o tempo. O número de variáveis a considerar pode ser reduzido e podem ser procuradas novas formas de representar a informação. Com a informação transformada, é preciso definir a técnica de mineração que mais se adapte ao processo do KDD. O modelo gerará uma série de padrões, que serão interpretados e avaliados, para finalmente serem convertidos em conhecimento, como apresenta a Figura 4. Este novo conhecimento servirá para executar ações com base na informação.

FIGURA 4 - PROCESSO DO KDD



Fonte: Adaptada de (FAYYAD; PIATETSKY-SHAPIRO; SMYTH, 1996)

### 2.3 Técnicas de Visualização de Informação

As técnicas de visualização de informação buscam aproveitar a capacidade dos humanos em extrair padrões a partir de imagens e ajudá-los a compreender de maneira mais rápida esses padrões. Na busca de extrair conhecimento a partir de grandes coleções de informação, essas técnicas permitem construir estruturas

mentais, ou cognitivas, para um domínio específico (CHEN, 2002). Elas também podem ressaltar características importantes de um conjunto de informações, transformando dados abstratos em formas visuais. De fato, as técnicas de visualização de informação não definem apenas métodos gráficos de apresentação das informações, mas são ferramentas criadas para apoiar a representação e a interação com os dados e, assim, dar apoio ao KDD (CRAFT; CAIRNS, 2008).

Três aspectos fundamentais caracterizam estas técnicas: (1) a estrutura; (2) a representação e (3) a interação da informação, com atenção especial às metáforas visuais, aos algoritmos de classificação, à distribuição da informação e à transformação interativa da informação visualizada, de acordo com os critérios estabelecidos pelos usuários.

Os dados manipulados pelas técnicas de visualização podem ser expressos de forma quantitativa e/ou qualitativa (BARROS; PRATES; MELO-MINARDI, 2014).

Os dados quantitativos são definidos como valores baseados em dados numéricos que representam algum tipo de informação, tais como: itens de escala Likert, coordenadas geográficas, codificações quantificadas, etc. (BARROS; PRATES; MELO-MINARDI, 2014).

Os dados qualitativos possuem um escopo mais amplo e podem estar na forma de entrevistas transcritas, gravações de conversas, imagens, vídeos, desenhos que capturam eventos, processos e resultados (BARROS; PRATES; MELO-MINARDI, 2014). Eles podem ser trabalhados e representados de maneira gráfica, tornando os textos utilizáveis, interativos e mais facilmente perceptíveis para a análise visual.

Segundo (BARROS; PRATES; MELO-MINARDI, 2014) não existem abordagens puramente qualitativas. Em sua análise, os autores concluem que as abordagens ou são quantitativas ou utilizam uma abordagem mista, unindo uma análise quantitativa com uma qualitativa. Em (GASPARINI; KIMURA; PIMENTA; ALEGRE, 2013), por exemplo, são utilizadas técnicas como Gráfico de Barras, Gráfico Circular, Nuvem de Termos, Mapas entre outras, para fazer uma abordagem mista de dados obtidos a partir das publicações do evento IHC (Simpósio Brasileiro

de Fatores Humanos em Sistemas Computacionais) entre 1998 e 2013. A análise foi baseada na exploração visual da base de informações dos artigos publicados no evento, com o objetivo de extrair conhecimento sobre a evolução de tópicos, autores e instituições centrais, além de desafios e direções de pesquisa mais importantes no período considerado.

Ao analisar os trabalhos de Bhowmick (2006) e Henderson e Segal (2013), ambos associados à análise de produção científica de um grupo de pesquisadores, Barros, Prates e Melo-Minardi (2014) afirmam que a técnica de visualização nuvem de termos é perfeitamente adequada para apresentar dados qualitativos com um nível de complexidade simples. “Ela é utilizada tanto qualitativamente através da análise do conteúdo dos termos, ou seja, do significado transmitido pelas palavras, quanto quantitativamente através da análise das frequências de aparição dos termos no texto” (BARROS; PRATES; MELO-MINARDI, 2014).

A abordagem mista aliada a formas gráficas adequadas para representar dados textuais utilizadas nestes últimos trabalhos citados é similar, em alguns aspectos, às técnicas de visualização propostas nessa dissertação.

### 2.3.1 Visualização de Informação em Nuvem de Palavras

A quantidade de informações acessadas pelos usuários cresce a cada dia, sendo que grande parte desta informação está em formato de texto (NUALART; PÉREZ; WHITELAW, 2014). Um texto pode ser definido como um dado sequencial, unidimensional e não estruturado. Apesar de ser um conteúdo digital relativamente simples de ser produzido, distribuído e processado, apenas recentemente os textos desperta maior interesse dos pesquisadores da área de visualização de informação.

Há menos de uma década, em 2010, os pesquisadores F. Viegas e M. Wattenberg relataram em uma entrevista a J. Heer, que uma das coisas que achavam realmente promissoras era a visualização de textos (HEER, 2010). Até aquela época, o texto vinha sendo ignorado em termos de ferramentas de visualização de informações, mais direcionadas à visualização de base de dados, apesar de uma grande quantidade de informação existir em formato de texto puro e não estruturado (HEER, 2010). De fato, (VIÉGAS; WATTENBERG, 2008) e (CUI;

WU; LIU; WEI; ZHOU; QU, 2010) observaram que a demanda do uso de nuvens de palavras indicava uma importante necessidade dos usuários: a visualização de informações contidas em textos não estruturados.

Uma Nuvem de Palavras (*Word Cloud*) ou Nuvem de Tags (*Tag Cloud*) é uma combinação de vários tamanhos diferentes de fonte em uma única visualização. A nuvem caracteriza uma das tentativas para melhorar a visualização das informações de um texto. Os primeiros indícios foram reportados há cerca de 90 anos, na era do construtivismo soviético (VIÉGAS; WATTENBERG, 2008).

Uma nuvem de palavras fornece um elemento gráfico tipográfico atraente que apresenta uma característica que a distingue das demais: o texto não precisa ser estruturado, sendo possível compreender as informações com uma simples visão geral. Esta forma de representação visual serve principalmente para identificar as palavras mais utilizadas em um texto, conhecer os conceitos enfatizados e para analisar a densidade dessas palavras-chave identificadas. Além do estilo visual, uma nuvem de palavras normalmente tem uma finalidade específica: apresentar uma descrição visual de uma coleção de dados textuais (VIÉGAS; WATTENBERG, 2008) (MCNAUGHT; LAM, 2010).

A nuvem de palavras pode ser utilizada para fins analíticos, comunicar padrões de textos e apresentar características das pessoas individualmente. Essa técnica é bem flexível, podendo processar grandes quantidades de texto, qualquer que seja o tema, e produzir rapidamente uma visão geral dos termos mais relevantes. Por exemplo, ela pode ser utilizada em periódicos e revistas para apresentar debates políticos de candidatos presidenciais ou para retratar os comentários dos cidadãos sobre as reformas de saúde, etc. (VIEGAS; WATTENBERG; FEINBERG, 2009). Além disso, pode ser adaptada em análises formais ou informais.

### 2.3.2 Visualização em Nuvem de Palavras Animadas

Enquanto uma nuvem de palavras trabalha com uma combinação de vários tamanhos diferentes de fonte em uma única visualização de um texto, sua versão

animada apresenta as informações do texto de acordo com algum atributo temporal (ordinal) dos dados, de maneira dinâmica e diferente. Neste sentido, ela se torna uma técnica atraente para apresentar, por exemplo, informações sobre a evolução de temas de pesquisa com o tempo ou ainda, detectar um tema de pesquisa que se mantém consistente durante um intervalo considerável de tempo, indicando a consolidação deste tema para uma área de pesquisa. Em redes sociais, elas podem indicar a tendência do uso de certos termos por uma comunidade atrelada a algum evento e pelos conteúdos produzidos e compartilhados pelos usuários finais por meio de tweets (Twitter) e posts (Facebook), por exemplo analisar sentimentos dos usuários no Twitter (ARAÚJO, 2013) ou em outras redes sociais (FELDMAN, 2013).

Neste trabalho, propõe-se a aplicação da metáfora da Nuvem de Palavras Animadas para apresentar informações relacionadas à produção de um grupo de pesquisadores, tendo como base a variação de frequência das palavras-chave dos títulos das publicações produzidas por eles.

Construir essa Nuvem de Palavras Animadas apresenta uma série de desafios, dentre os quais podem ser destacados: (1) entender quais informações podem ser acompanhadas durante a animação; (2) entender o significado que é atribuído às mudanças observadas no intervalo de tempo apresentado e (3) interpretar a dinâmica das informações apresentadas.

A técnica de visualização utiliza as diferenças de tamanho do texto nas nuvens geradas para evidenciar a quantidade de trabalhos produzidos sobre um determinado tema em determinado período e a animação destes tamanhos para evidenciar o momento do surgimento, a manutenção da relevância ou o esgotamento de um tema nas pesquisas do grupo.

### 2.3.3 Visualização em Grafos

A visualização de informação em grafos pode informar e revelar as relações e padrões da informação, sem saber nada sobre os dados subjacentes, por meio de estruturas simples, elegantes e diretas (OLMEDA, 2014).



Os grafos permitem expressar visualmente, de maneira muito simples e eficaz, os relacionamentos que existem entre os elementos de vários tipos (DÜRSTELER; 2004). Eles são especialmente úteis na representação visual de estruturas de dados e redes hierárquicas (MONTERO, 2006).

Os relacionamentos podem ser simples ou complexos, diretos ou bidirecionais, ponderados ou não ponderados, ou com certos graus de incerteza. Eles podem fornecer mais informações além das contidas nos registros dos bancos de dados tomados um a um (OLMEDA, 2014). Há informações adicionais nas conexões. Com um só olhar, podem ser observados os nós que atraem um número maior de ligações, quais são agrupados e formam diferentes subgrupos e ainda, os nós isolados, que não têm conexões.

Neste trabalho, propõe-se a aplicação da visualização de informação em grafos para apresentar o relacionamento das produções bibliográficas produzidas por grupos de pesquisadores. Também para observar quais são os temas que mais relevância têm nesses grupos, quantas publicações são associadas ao tema eles e quais são os tipos (conferência, periódico etc.) de publicação produzidas.

#### 2.3.4 Visualização em Gráfico de Barras

Os gráficos de barras são uma das maneiras mais comuns para se visualizar dados. A comparação de informação é rápida, revelando pontos altos e baixos em um olhar (HARDIN; HOM; PEREZ; WILLIAMS, 2011). Os gráficos de barras são uma forma de representar graficamente informação ou um conjunto de valores. Eles consistem em barras retangulares de comprimentos proporcionais aos valores representados. São especialmente eficazes quando se tem dados numéricos que se dividem ordenadamente em diferentes categorias, permitindo identificar tendências no conjunto de dados de forma bastante rápida.

Neste trabalho, os gráficos de barras são bastante úteis porque permitem comparar o trabalho dos pesquisadores. Por exemplo, são utilizados para comparar o número de produções bibliográficas com o número de trabalhos orientados de um

pesquisador individualmente em um ano específico. Além de visualizar eficazmente a informação numérica dos pesquisadores de um grupo.

### 2.3.5 Visualização em Gráfico Circular

O gráfico circular consiste em um círculo dividido em partes, cujas dimensões são proporcionais a algum dos atributos dos dados a serem visualizados. O círculo é dividido em tantos setores quantos forem as classes de valores a serem representadas. A dimensão de cada setor deve ser proporcional à frequência do valor correspondente. O objetivo é que seja possível obter as informações associadas aos dados representados no círculo a partir das medidas de cada uma de suas partes. Os gráficos de círculos são especialmente indicados para mostrar proporções ou porcentagens de informações (HARDIN; HOM; PEREZ; WILLIAMS, 2011).

Neste trabalho, os Gráficos Circulares são utilizados para representar diversas questões. Primeiramente, são utilizados para conhecer e comparar a quantidade das Produções Bibliográficas dos professores em um grupo de pesquisa, permite também conhecer a porcentagem e representar graficamente qual professor publicou mais em um ano especificamente. Também, permite visualizar de forma individual qual é o tema majoritariamente pesquisado por um professor.

## 2.4 Considerações do capítulo

A visualização de informação tem três grandes objetivos: descrever, explorar e analisar a informação, com o objetivo de descobrir e amplificar o conhecimento. Obter *insight* das informações é seu principal fim.

A visualização de informação deve ampliar a cognição a partir de representações visuais das estruturas contidas na informação e com a interação dela. Os modelos apresentados nesse capítulo buscam representar elementos fundamentais da informação, sua transformação, a interação e a percepção, para conseguir uma visualização adequada do conjunto de informações. Assim também, as transformações são feitas a partir da informação para estabelecer estruturas,

relações, conceitos e uma redução dimensional, utilizando uma linguagem gráfica previamente selecionada para obter novas perspectivas e novos conhecimentos.

É importante considerar a complexidade, a percepção do usuário, o domínio onde é precisada a ferramenta e o contexto de trabalho no momento de desenvolver as técnicas de visualização de informação.

Esse capítulo reuniu a informação documental para criar um conhecimento aprofundado da teoria que dá significado à pesquisa. A partir das teorias existentes do objeto de estudo (visualização de informação), é possível gerar novos conhecimentos. O capítulo corresponde à etapa do processo de pesquisa, onde é estabelecida a teoria que é seguida como modelo da realidade que se está pesquisando e que dá a sustentação teórica ao estudo. Nele, foi feita uma análise crítica de diferentes tendências ou teorias enunciadas e também a revisão da literatura sob a base dos resultados de outras pesquisas. O referencial teórico tem o objetivo de dar à pesquisa um sistema coordenado e coerente de conceitos que permitem abordar o problema e que deste dependerá o resultado do trabalho.

### **3 TRABALHOS RELACIONADOS**

### 3 TRABALHOS RELACIONADOS

Neste capítulo serão discutidos os trabalhos relacionados ao tema principal, ferramentas para visualização e extração de conhecimentos de um conjunto de informações.

Diversas ferramentas ou recursos de visualização estão disponíveis para a análise e visualização de informação, as quais possibilitam desde a obtenção de informações simples, até a identificação de padrões e tendências que levam a *insights* significativos (VIÉGAS; WATTENBERG, 2008).

O capítulo divide os trabalhos relacionados em duas partes bastante distintas. A primeira trata dos trabalhos direcionados à visualização de informação qualitativa que serviram de base para o projeto e implementação do sistema proposto nessa dissertação. A segunda parte apresenta trabalhos relacionados à visualização de trabalhos científicos produzidos por diferentes comunidades brasileiras ligadas à computação e, portanto, com alguma similaridade à proposta dessa dissertação.

#### 3.1 Many Eyes

Uma das ferramentas mais conhecidas para visualização de dados é o Many Eyes<sup>1</sup> (VIEGAS *et al.*, 2007). Trata-se de uma ferramenta colaborativa, lançada em 2007 e desenvolvida por um grupo de pesquisa da empresa IBM, criado por Martin Wattenberg no ano 2004. Ela é uma ferramenta de uso simples que permite que os usuários façam o upload de dados e criem diversas visualizações a partir deles. Um dos seus principais objetivos é tentar transformar a informação em gráficos, de modo que certos conhecimentos sejam disponibilizados para todas as pessoas. Uma

---

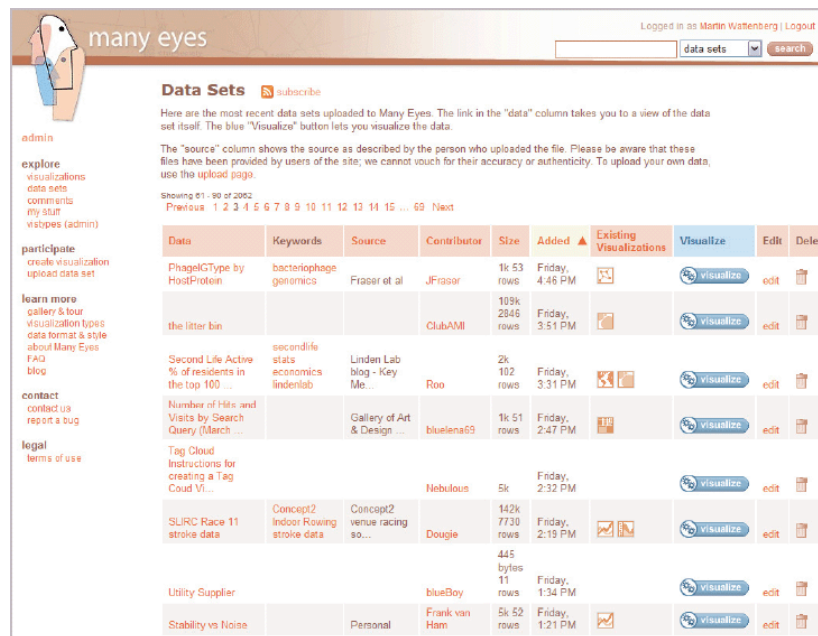
<sup>1</sup> <http://manyeyes.alphaworks.ibm.com/manyeyes/>

questão importante para o uso do Many Eyes é preparar bem os dados e ter clareza do tipo de visualização que o usuário quer obter.

Entre as muitas técnicas de visualização de informação que o Many Eyes disponibiliza está a Tag Cloud. Essa forma de visualização foi incluída para atender aos usuários que tinham dados não estruturados como entrada e queriam contar a frequência de palavras (VIEGAS *et al.*, 2007). Os criadores notaram que as nuvens de palavras atraíam um novo tipo de usuário, mais interessados em dados textuais que em dados numéricos.

Uma importante contribuição deste trabalho foi a de oferecer grande relevância e suporte para a informação de tipo texto, permitindo criar diversas técnicas de visualização de informação com esses dados. Porém, não foi possível utilizar esta ferramenta diretamente nessa dissertação porque ela não é um *software open source*, não é adaptável e também porque a empresa IBM decidiu encerrar o projeto em meados de 2015.

FIGURA 5 - UMA DAS INTERFACES DO MANY EYES



The screenshot shows the Many Eyes website interface. At the top, there is a navigation bar with the 'many eyes' logo, a user login status 'Logged in as Martin Wattenberg | Logout', and a search bar. Below the navigation bar, there is a 'Data Sets' section with a 'subscribe' button. The main content area displays a table of data sets with columns for Data, Keywords, Source, Contributor, Size, Added, Existing Visualizations, Visualize, Edit, and Delete. The table lists several data sets, including 'PhageGType by HostProtein', 'the litter bin', 'Second Life Active % of residents in the top 100', 'Number of Hits and Visits by Search Query (March)', 'Tag Cloud Instructions for creating a Tag Cloud Vi...', 'SURC Race 11 stroke data', 'Utility Supplier', and 'Stability vs Noise'.

Data	Keywords	Source	Contributor	Size	Added	Existing Visualizations	Visualize	Edit	Delete
PhageGType by HostProtein	bacteriophage genomics	Fraser et al	JFraser	1k 53 rows	Friday, 4:46 PM		<a href="#">visualize</a>	<a href="#">edit</a>	
the litter bin			ClubAMI	109k 2846 rows	Friday, 3:51 PM		<a href="#">visualize</a>	<a href="#">edit</a>	
Second Life Active % of residents in the top 100 ...	secondlife stats economics lindentab	Linden Lab blog - Key Me...	Roo	2k 102 rows	Friday, 3:31 PM		<a href="#">visualize</a>	<a href="#">edit</a>	
Number of Hits and Visits by Search Query (March ...		Gallery of Art & Design ...	bluelena69	1k 51 rows	Friday, 2:47 PM		<a href="#">visualize</a>	<a href="#">edit</a>	
Tag Cloud Instructions for creating a Tag Cloud Vi...			Nebulous	5k	Friday, 2:32 PM		<a href="#">visualize</a>	<a href="#">edit</a>	
SURC Race 11 stroke data	Concept2 Indoor Rowing stroke data	Concept2 venue racing so...	Dougle	142k 7730 rows	Friday, 2:19 PM		<a href="#">visualize</a>	<a href="#">edit</a>	
Utility Supplier			blueBoy	445 bytes 11 rows	Friday, 1:34 PM		<a href="#">visualize</a>	<a href="#">edit</a>	
Stability vs Noise		Personal	Frank van Ham	5k 52 rows	Friday, 1:21 PM		<a href="#">visualize</a>	<a href="#">edit</a>	

Fonte: Adaptada de (VIEGAS *et al.*, 2007)

### 3.2 Zingchart

ZingChart<sup>2</sup> é uma biblioteca *opensource* desenvolvida em JavaScript para o HTML5 que faz a criação de gráficos interativos com o objetivo de apoiar a visualização de informação por meio de diversas técnicas. Essa biblioteca tem como objetivo ser flexível, interativa, rápida e escalável. O ZingChart permite criar diferentes tipos de visualizações de informação a partir de um conjunto de dados.

ZingChart surgiu a partir do desenvolvimento de uma análise Web profunda, no que se encontrou a necessidade de ter melhores características em gráficos para criar visualizações de informação. Como resultado, o projeto ZingChart nasceu no ano 2008, sendo lançado na Web um ano depois.

O projeto desenvolvido neste trabalho está baseado na ferramenta ZingChart por ser muito completa e versátil de utilizar, ela permitiu criar algumas técnicas de visualização de informação a partir de um conjunto de dados fornecidos em formato de tipo texto. O ZingChart foi utilizado para a filtragem dos dados de entrada no trabalho.

### 3.3 WordSwarm

WordSwarm<sup>3</sup> é um projeto desenvolvido em Python, de código aberto, licenciado com GPLv3 e disponível para utilização. Ele foi originalmente concebido para analisar as tendências de variação na área financeira, interesses científicos e nomes populares em um intervalo de tempo. Entretanto, o projeto, desenvolvido por Michael Kane, acabou sendo usado com sucesso para gerar visualizações dinâmicas de temas de interesse a partir da frequência das palavras presentes em periódicos. Por ser um projeto interessante, se procurou maiores informações do projeto diretamente do autor para a utilização e apoio no desenvolvimento deste

---

<sup>2</sup> <https://www.zingchart.com/about-us/>

<sup>3</sup> [http://www.thisismikekane.com/proj\\_wordswarm.php?filter=](http://www.thisismikekane.com/proj_wordswarm.php?filter=)

trabalho. Foram recebidas algumas sugestões do autor e a técnica de visualização de nuvens de palavras animadas foi baseada neste projeto.

### **3.4 Pygame**

Pygame<sup>4</sup> consiste em um conjunto de módulos Python, os quais permitem a criação de aplicações multimídia ou interfaces gráficas com animações. É também muitas vezes usado para criar jogos em duas dimensões de uma forma simples. A animação criada na técnica de visualização de nuvens de palavras animadas usada na implementação descrita no próximo capítulo foi baseada nos módulos Pygame.

As seções seguintes do texto irão apresentar alguns trabalhos recentes com foco na análise da produção científica de pesquisadores nas conferências nacionais apoiadas pela Sociedade Brasileira de Computação. Deve-se salientar, porém, que os trabalhos possuem um viés mais ligado à análise das relações entre pesquisadores (análise de redes sociais) do que na qualificação das publicações.

### **3.5 Utilizando títulos de artigos científicos na construção de redes semânticas para caracterizar áreas de pesquisa**

O trabalho de Cruz (2015) apresentou uma técnica que caracteriza áreas de pesquisa científica através da construção de redes semânticas variáveis no tempo. Após os devidos tratamentos, as redes semânticas foram construídas tomando por base a concorrência de palavras-chave nos títulos de artigos científicos, que servem para representar a evolução dos temas de pesquisa abordados pelos meios de publicação ao longo do tempo.

Os resultados deste trabalho apresentam que os usos de redes semânticas baseadas em títulos, geralmente, auxiliam no entendimento da integração das

---

<sup>4</sup> <http://www.pygame.org/hifi.html>



palavras-chave. Além disso, o uso de técnicas de visualização de informação juntamente com exploração visual manual permitem e facilitam a combinação de elementos para o entendimento e a avaliação das estruturas das redes semânticas dos títulos dos artigos científicos.

Uma importante relação deste trabalho com a pesquisa, é que baseado em títulos e palavras-chave de artigos científicos se construíram redes semânticas para a extração de conhecimento, neste caso, para caracterizar as áreas de pesquisa. O trabalho utiliza a ferramenta de visualização de informação Gephi<sup>5</sup> (BASTIAN; SEBASTIEN; MATHIEU, 2009), para apresentar os resultados. Gephi é uma ferramenta de visualização de informação que trabalha só com técnicas de visualização de redes e grafos.

### **3.6 Análise da rede de colaboração do SBRC: As primeiras 30 edições**

O artigo de Maia et al. (2012) apresentou o estudo da rede de colaboração entre os autores do Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos (SBRC), coletando os dados bibliográficos das trinta edições do evento. Eles também construíram a rede de coautoria do SBRC e analisaram tanto as características estruturais quanto a sua evolução ao longo da história.

Neste trabalho também é possível observar a informação e conhecimento que são extraídos dos dados que poderiam ser simples, mas se são apresentados e analisados de uma forma ótima podem dizer mais que ser só um título de um artigo científico.

---

<sup>5</sup> <https://gephi.org/>

### 3.7 Visualizando 15 anos do IHC

Gasparini *et al.* (2013) apresentam a análise da produção bibliográfica de 15 anos do Simpósio Brasileiro sobre Fatores Humanos em Sistemas Computacionais (IHC). Essa análise foi baseada na exploração visual dos dados básicos como ano, título, palavras-chave e outros, para identificar a evolução dos temas pesquisados, os autores, as instituições principais que publicam e outras tendências importantes.

Da mesma forma que esta dissertação, o trabalho utilizou majoritariamente a técnica de visualização de informação nuvens de palavras para apresentar a evolução dos temas baseados principalmente nas palavras-chave.

Para a análise e interpretação dos resultados foram realizados filtros na base de dados, de acordo com as especificações de cada análise.

### 3.8 WebMedia, 18 anos: Quem somos e o que temos feito nas últimas duas décadas?

Duarte *et al.* (2014) publicaram um artigo onde analisaram os artigos científicos de 18 edições do congresso WebMedia, com o objetivo de apresentar como os temas de pesquisa têm evoluído no tempo. A inovação deste artigo foi que eles utilizaram as técnicas de análise de redes sociais para identificar grupos de busca, *clusters* e os temas apresentados nos eventos WebMedia nas décadas de 1990 e 2000. Para a extração da informação relevante do que queria ser apresentado, os autores utilizaram os campos Título do artigo, Ano de publicação, Lista de autores, campos que também foram considerados nesta pesquisa.

### 3.9 Considerações do Capítulo

Atualmente, os recursos de visualização disponíveis para obtenção, análise dos dados e criação das técnicas de visualização de informação ainda não estão devidamente integradas. Portanto, faz-se necessário, a utilização isolada de algumas ferramentas para obter os resultados esperados e integrar as técnicas de visualização de informação criadas em um *site*.

A partir da análise dos principais estudos relacionados diretamente ao tema trabalhado, é possível fazer uma diferença. Este trabalho não pretende estudar e analisar as produções feitas ao longo do tempo pelos professores que pertencem a um programa de estudo de uma universidade. Porém, algumas das diversas técnicas de visualização de informação utilizadas nos trabalhos relacionados foram selecionadas para ser adaptadas na dissertação. Técnicas de visualização de informação, como a nuvem de palavras, foram estendidas e modificadas no trabalho, acrescentando dimensão temporal e dinamismo nelas.

A análise das Produções Bibliográficas permitirá que se tenha como saída, também, informação quantitativa e não só informação qualitativa, que fará possível a extração de conhecimentos das visualizações.

Um diferencial importante que o trabalho apresenta é a categorização de um trabalho, a partir de suas palavras-chave, de acordo com as categorias atuais propostas pela ACM<sup>6</sup>. Também será possível observar se o termo utilizado como palavra-chave é recomendado pela ACM.

---

<sup>6</sup> <http://dl.acm.org/>



## 4 A PROPOSTA

Após serem estudadas algumas conhecidas técnicas de visualização de informação utilizadas no campo científico, assim como vários conhecidos recursos de visualização disponíveis e abertos a pesquisadores deste domínio, foi selecionado um cenário real de aplicação da proposta desta dissertação. Dentro desta perspectiva, um estudo de caso envolvendo uma base de informações de pesquisadores ligados ao Programa de Pós-graduação de Informática (PPGI) da UFES foi construído e usado para avaliação da proposta.

Os critérios para escolher as técnicas e ferramentas aplicadas ao cenário escolhido foram: (1) as técnicas deveriam estar no estado da arte e poderiam ser modificadas ou estendidas; (2) elas deveriam poder ser utilizadas em diferentes cenários de aplicação e com diferentes tipos de dados, especialmente os dados qualitativos (texto) e (3) elas deveriam prover uma exploração interativa das informações e levar em conta a dimensão temporal das informações.

A partir da observação e extensão de trabalhos anteriores, algumas técnicas de visualização de informação foram então propostas e implementadas: A Nuvem de Palavras Animadas Interativa; a Rede das Produções Bibliográficas (Grafos); Árvore que permite categorizar uma publicação de acordo com as categorias da ACM. Estas técnicas serão detalhadas no decorrer das próximas seções deste capítulo.

### 4.1 Cenário real da aplicação

A Plataforma Lattes do CNPq<sup>7</sup> tornou-se o padrão nacional no registro do percurso acadêmico de estudantes e pesquisadores do Brasil. Essa plataforma contém a informação acadêmica dos CV de professores, grupos de pesquisa e de instituições em um único sistema de informação.

---

<sup>7</sup> <http://lattes.cnpq.br>

O cenário real da aplicação utiliza as informações sobre as Produções Bibliográficas dos grupos de pesquisadores do PPGI-UFES, que foram obtidas diretamente da plataforma. Apesar dos dados serem restritos a estes pesquisadores, por conta do interesse específico deste trabalho, o mesmo processo pode ser replicado a outras bases de informações similares à utilizada, seja em outros programas ou áreas de conhecimento para as quais os dados estejam disponíveis via a Plataforma Lattes.

#### **4.2 O “VI-Project”**

Com o objetivo de modelar um estudo de caso, foi desenvolvido um *site* para avaliar as técnicas de visualização de informação num cenário real. O VI-Project é um *site* que apresenta diversas destas técnicas sobre um conjunto de informações acadêmicas de professores e grupos de pesquisadores de um Programa de Pós-graduação. O acesso ao VI-Project pode ser feito a partir do endereço [http://200.137.66.2/cvfront/Web/app\\_dev.php/login](http://200.137.66.2/cvfront/Web/app_dev.php/login) e os passos para a sua implantação estão no Apêndice 1.

No *site* é possível observar com detalhe os temas de pesquisa relevantes tanto para os professores que formam um grupo de pesquisa como os temas relevantes para um professor de forma individual, sendo que essa informação poderia se apresentar em um determinado ano ou em um intervalo. A apresentação dessa informação foi baseada em algumas técnicas de visualização descritas no início do capítulo (Nuvens de Palavras Animadas, Gráficos de Barra e Circular, Grafos, etc.), porém, essas técnicas foram estendidas para que se tornassem dinâmicas, interativas e apresentassem um volume maior de informação (conhecimento) dos que apresenta comumente.

Também, nos dados obtidos dos CV Lattes dos professores se apresenta a informação de seus Trabalhos Orientados. No desenvolvimento do *site*, se encontrou uma ótima forma para apresentar essa informação. Contrastar a quantidade e tipo de Trabalhos Orientados com a quantidade e tipo de Produções Bibliográficas por meio de um gráfico de barras poderia oferecer algumas evidências sobre as relações entre produções discentes e docentes no programa. Por exemplo, seria possível

visualizar se um professor possui maior quantidade de Trabalhos Orientados ou de Produções Bibliográficas, qual tipo de Publicação Bibliográfica desenvolve mais ou o tipo de orientação mais realizada por ele.

Mostrar uma rede de nós (Grafos) das Produções Bibliográficas dos Grupos de Pesquisa busca evidenciar os temas que se desenvolvem. Se um nó principal tem relação com muitos nós das Produções Bibliográficas, indica que essa é a área de atuação principal do grupo. Também é possível evidenciar aquelas áreas de atuação que não são tão relevantes ou aquelas Produções Bibliográficas que não tem relação com alguma área de atuação. Grafos são normalmente utilizados para representar estruturas de dados ou estruturas hierárquicas, porém, nesse trabalho foi modificada e estendida para que pudesse representar a informação que se requeria.

Apresentar as Produções Bibliográficas com diferentes gráficos e com cálculos estatísticos poderia ajudar a identificar o docente líder de um Grupo de Pesquisa, qual é o tipo de trabalho mais publicado do Grupo de Pesquisa, ou o total ou a média das Produções Bibliográficas em um ano. Para representar cálculos de uma forma gráfica ótima e dinâmica se fez uso dos gráficos circulares que promovem uma visão geral rápida do que se quer representar.

O VI-Project conta com cinco módulos que apresentam diferentes informações: (1) Professores; (2) Grupos de Pesquisa; (3) Nuvens de Palavras; (4) Administração de Usuários e (5) Termos da ACM.

### **4.3 Metodologia de Análise e Desenho do Sistema**

A metodologia de análise e desenho do sistema proposta por (KENDALL; KENDALL, 2011) serviu de base para o desenvolvimento. A metodologia possui sete fases: (1) Identificação de problemas, oportunidades e objetivos; (2) Determinação dos requisitos; (3) Análise das necessidades; (4) Desenho do sistema; (5) Desenvolvimento do sistema; (6) Testes e (7) Implementação.

Da mesma forma, foram criados Diagramas de Casos de uso UML para descrever as funcionalidades do sistema do ponto de vista do usuário, e o Diagrama da Estrutura do banco de dados.

#### 4.3.1 Identificar problemas, oportunidades e objetivos

Nesta primeira fase se identificou um problema importante: Os dados dos professores necessários para criar as técnicas de visualização se encontravam em uma plataforma na qual não era muito simples manipular e extrair os dados (seção 4.5.1).

Também, na revisão da literatura, foram encontradas técnicas utilizadas para visualizar a informação de artigos científicos, essas técnicas eram susceptíveis a melhoras neste projeto. Por exemplo, as nuvens de palavras podiam evoluir e ser apresentadas de forma dinâmica. Também, podiam se utilizar os gráficos de barra e circulares em apoio à visualização de dados de tipo texto.

#### 4.3.2 Determinação dos requisitos

Nesta segunda fase se determinou o processo para a migração da informação e o armazenamento no banco de dados para sua manipulação. Também, foi possível compreender a informação necessária para a criação das técnicas de visualização e o que se queria realmente apresentar. Isso permitiu determinar quais eram especificamente as técnicas de visualização que se pretendiam criar.

A partir da análise dos trabalhos relacionados ao tema, disponibilidade de ferramentas, determinou-se que as técnicas de visualização a serem usadas no trabalho seriam: nuvem de palavras animadas acrescentado dimensão temporal e dinamismo, diagramas circulares e de barra, diagramas de grafo ou de rede e diagrama de árvore.



#### 4.3.3 Análise das necessidades

Nesta fase, foram analisadas as necessidades próprias do *site* e do banco de dados. Aqui foram criados os Diagramas de Pacotes UML para observar os processos do *site* claramente. Além disso, foi criado o esquema do banco de dados para continuar com a estrutura dos dados obtidos da plataforma CV Lattes.

#### 4.3.4 Desenho do sistema

Esta fase vai ser detalhada na seção 4.5 do capítulo.

#### 4.3.5 Desenvolvimento do sistema

Esta fase vai ser detalhada na seção 4.5.2 do capítulo.

#### 4.3.6 Teste do sistema

Antes de utilizar o *site* final, devem realizar-se os testes devidos para descobrir possíveis problemas no funcionamento. Foram realizados tanto os testes na conexão com o banco de dados, quanto os testes para apresentar a informação nas técnicas de visualização desenvolvidas. Os testes serão detalhados no capítulo dedicado ao estudo de caso desta dissertação.

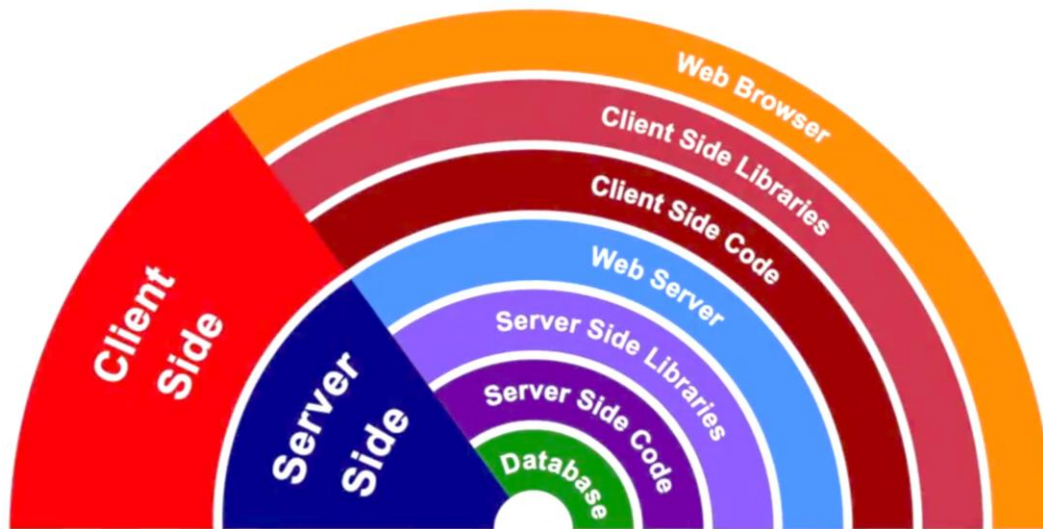
#### 4.3.7 Implementação do sistema

Na última fase da metodologia se hospedou o *site* em uma máquina virtual disponibilizada pelo Departamento de Informática da UFES, para seu completo acesso.

### 4.4 Camadas do sistema Web implementado

A Figura 6 ilustra as camadas presentes em um sistema Web, as quais são divididas em dois grupos: lado do cliente e lado do servidor.

FIGURA 6 - CAMADAS DE UM SISTEMA WEB



Fonte: Adaptada de (ROSSITE, 2016)

A partir da arquitetura anterior, as tecnologias e linguagens utilizadas no sistema Web apresentado nessa dissertação foram os seguintes:

- Na camada mais externa do lado do cliente está o *navegador*, as bibliotecas (Zingchart, JQuery, Bootstrap) e as linguagens utilizadas (HTML, JavaScript, CSS).
- Na camada mais interna do lado do servidor encontrasse o servidor Web utilizado (Apache), os *Frameworks* também usados (Symfony, Silex), a linguagem empregada (PHP) e o banco de dados criado (MySQL).

#### 4.5 Detalhes sobre o desenho do sistema

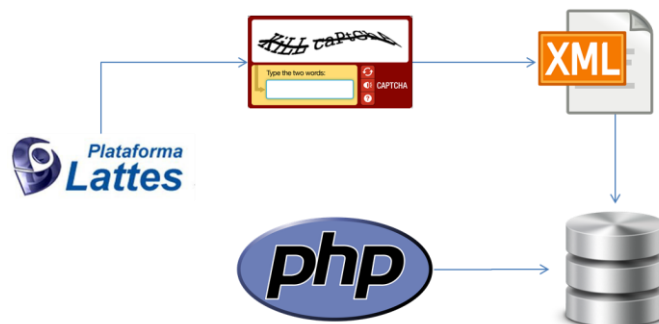
O desenho do sistema foi proposto em duas etapas. A primeira etapa foi a migração da informação para o banco de dados. A segunda etapa foi o processamento da informação para gerar as visualizações. Muitas ferramentas e métodos foram utilizados no processo do desenvolvimento, essas ferramentas estão descritas no Apêndice 2.

#### 4.5.1 Processo de Migração da Informação

As informações textuais utilizadas para a geração das visualizações consideradas neste trabalho foram obtidas diretamente da plataforma CNPq-Lattes. A ideia inicial foi fazer *Web Scraping* (HOGUE; KARGER, 2005) na base do Lattes, porém, a plataforma começou, a partir de 2015, a aplicar o controle de acesso com o uso de um *Captcha*. A solução encontrada, então, foi realizar o *download*, de forma manual e colaborativa, do arquivo XML de cada um dos CV dos pesquisadores pertencentes ao grupo a ser analisado.

Para a gestão e armazenamento das informações, foi desenvolvido um programa em PHP, a fim de gerenciar os dados obtidos (em XML) do Lattes, em um banco de dados MySQL. Esse programa interpreta a estrutura dos arquivos XML, coleta a informação selecionada, faz a conexão com a base de dados e armazena, nesse banco, os dados considerados. O processo de obtenção dos dados é apresentado na Figura 7.

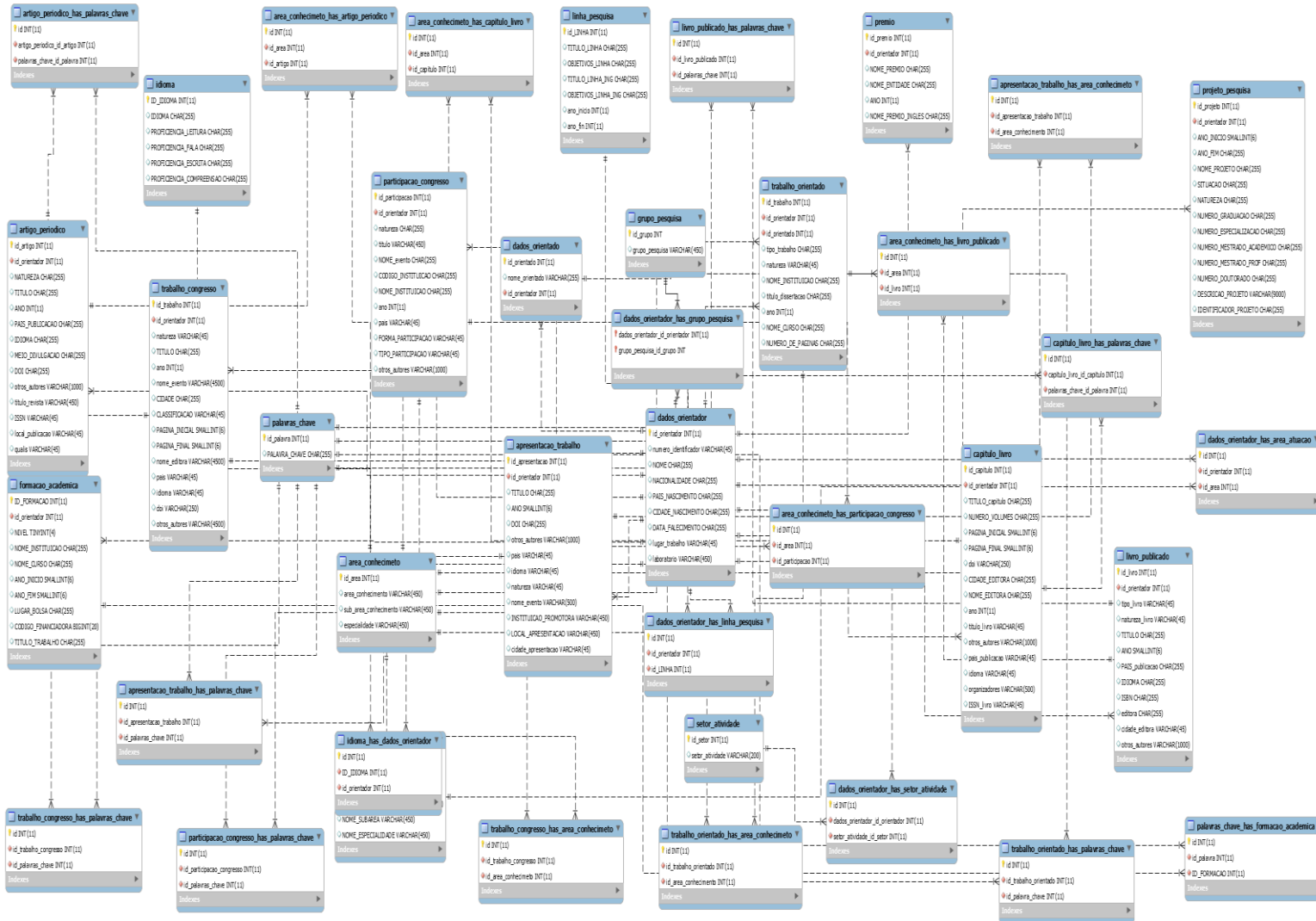
FIGURA 7 - PROCESSO DE OBTENÇÃO DOS DADOS A PARTIR DA PLATAFORMA LATTES



Fonte: Elaborado pelo autor

Para a criação do banco de dados MySQL, foi previamente desenhado o Modelo Entidade – Relacionamento (MER) que permitiu manter de forma normalizada a informação que continha os arquivos XML. A razão de usar esse modelo foi a persistência dos dados com a mesma estrutura dos arquivos XML. O esquema do banco de dados está na Figura 8.

FIGURA 8 - ESQUEMA DO BANCO DE DADOS



Fonte: Elaborado pelo autor

#### 4.5.2 Processo de Geração das Visualizações

O processo de geração das visualizações no sistema foi dividido em duas partes: o Processo de Geração das Visualizações Estatísticas e o Processo de Geração da Nuvem de Palavras Animadas. De fato, a maior parte das técnicas de visualização propostas nesse trabalho está associada ao primeiro processo.

##### 4.5.2.1 Visualizações Estatísticas

A operação de uma aplicação Web tradicional conta com um lado do cliente e um lado do servidor. O lado do cliente trata, principalmente, da apresentação dos dados para os usuários, usando tecnologias como HTML, CSS e JavaScript, quando eles acessam a Web por meio de um navegador. Enquanto isso, o lado do servidor está mais relacionado com a lógica de negócios, cálculos e o processamento dos dados da aplicação Web. Estes processos podem ser desenvolvidos em várias linguagens de programação.

Um padrão arquitetural bastante conhecido e utilizado para desenvolvimento cliente-servidor é o chamado Modelo-Visão-Controlador (*Model-View-Controller*) ou, simplesmente, MVC. O padrão divide uma aplicação em três camadas (o modelo, a visão e o controlador) e foi criado com o objetivo de reduzir o esforço do desenvolvimento, necessário na implantação de sistemas múltiplos e sincronizados com os mesmos dados, a partir da padronização do desenho dos aplicativos.

1. A camada de apresentação (Visão) trata das questões relacionadas com a interface do usuário final e a forma de apresentar os dados.
2. A camada da lógica de negócios (Controlador) está mais preocupada com a validação e o processamento de dados, com o objetivo de gerar informações para os usuários.
3. A camada de acesso aos dados (Modelo) lida com a persistência, o acesso e o armazenamento dos dados no banco de dados.

No entanto, a camada da lógica de negócios é geralmente realizada em um servidor de aplicativos Web, onde se pode usar qualquer linguagem de programação

para o desenvolvimento (PHP, Java, C#, etc.). A referida camada (lógica de negócios) recebe o pedido emitido pelo cliente (*request*), gerencia a comunicação entre o sistema de gestão do banco de dados para a troca de informação (camada de acesso aos dados) e encaminha os dados processados para a camada visão (camada de apresentação), permitindo, assim, a geração da visualização no lado do cliente utilizando as linguagens e bibliotecas do lado do cliente para retornar uma resposta (*response*).

O acesso aos dados armazenados baseia-se na resposta emitida pelo controlador através de um arquivo no formato JSON. É preciso ressaltar que o formato JSON tornou-se um dos formatos padrão<sup>8</sup> para a troca de dados entre as três camadas.

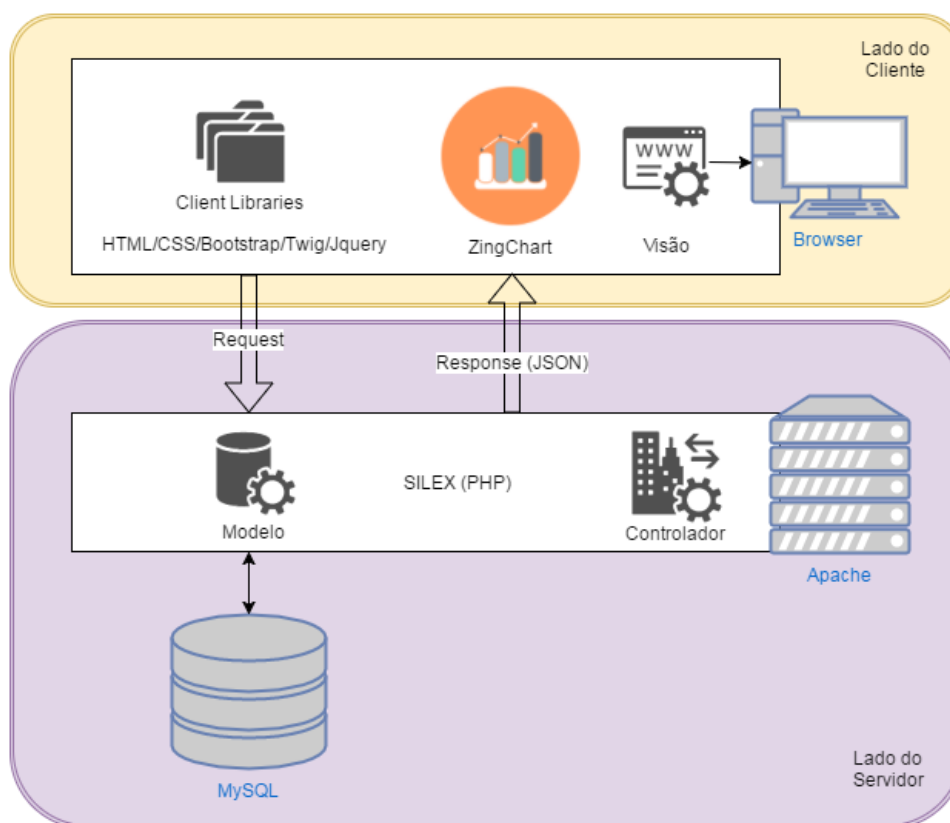
Especificamente, o sistema proposto cumpre com os seguintes passos, conforme mostrado na Figura 9.

1. Realiza-se uma solicitação HTTP através do navegador para a aplicação Web (da camada da Visão para a camada do Controlador) enviando por URL os parâmetros para fazer a consulta;
2. Com a solicitação se faz a chamada das bibliotecas do lado do Cliente para que o aplicativo comece a funcionar;
3. O Controlador direciona o fluxo para a camada do Modelo onde se faz a consulta ao banco de dados;
4. Os dados retornados pela camada do Modelo são enviados para a camada do Controlador, onde são processados e encaminhados para a camada da Visão no formato JSON;
5. Na camada da Visão, o motor de visualização interpreta os dados em formato JSON, gera o gráfico correspondente e volta a resposta (*response*) feitas pelas bibliotecas para o cliente (navegador).

---

<sup>8</sup> <http://www.json.org/json-pt.html>

FIGURA 9 - PROCESSO DE CRIAÇÃO DAS VISUALIZAÇÕES ESTADÍSTICAS



Fonte: Elaborado pelo autor

#### 4.5.2.2 Nuvem de Palavras Animadas

A biblioteca ZingChart foi utilizada para gerar uma Nuvem de Palavras principal, estática, que faz a filtragem da informação para cada grupo de pesquisa do programa de estudo por ano. Como saída, a biblioteca oferece uma imagem da Nuvem de Palavras, além de arquivos PDF e outros formatos. A ZingChart gera a quantidade de vezes que uma palavra aparece em uma publicação, indicando assim sua relevância. O processo também exclui as palavras cujas ocorrências devem ser desconsideradas (as chamadas *stop words*): *of, the, if, a, e, o, do, das, no, nas, etc.*

A partir destes dados e considerando as palavras que tem incidência maior ou igual a dois, foi possível gerar um arquivo CSV contendo as palavras da nuvem e o número de ocorrência de cada uma delas por ano. Este arquivo é utilizado como entrada para a segunda parte do processo proposto que é a animação das palavras na nuvem.

O componente responsável pela animação das palavras na solução proposta é baseado na plataforma WordSwarm. A WordSwarm foi criada para realizar um *Web Scraping* diretamente de um *Website* e extrair as informações relevantes, posteriormente gerando um arquivo CSV com a filtragem da informação. Como não foi preciso realizar esse processo, o código da aplicação original foi modificado para permitir que a entrada do programa, para fazer a criação da animação, fosse o arquivo CSV gerado anteriormente com o ZingChart e fosse utilizado só os últimos três passos da solução proposta WordSwarm como apresenta a Figura 10. Entre outras vantagens, esta estratégia permite reduzir bastante o tempo de processamento das informações e codificação dos vídeos com as animações, uma vez que o ZingChart faz todo o trabalho de filtragem das informações irrelevantes para a análise. Na solução original, todo esse trabalho anterior à geração da animação seria realizado pela plataforma WordSwarm.

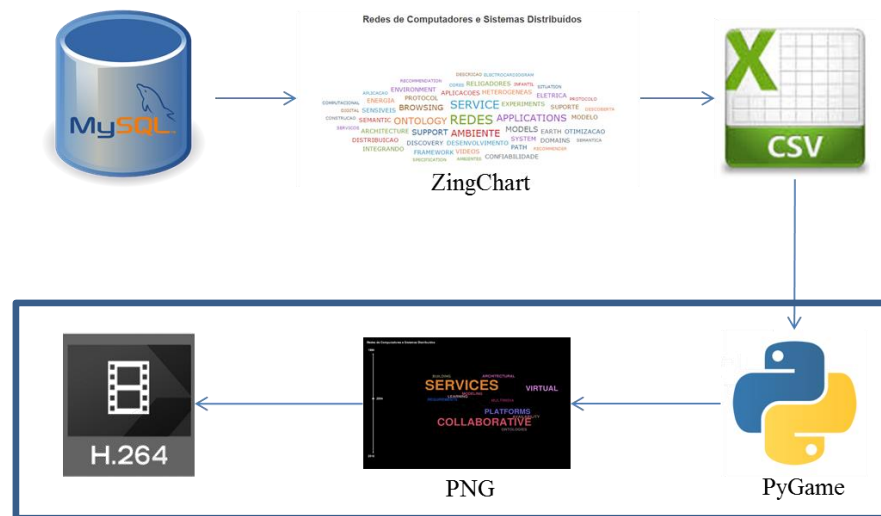
O arquivo CSV foi manipulado a partir da biblioteca PyGame, que é encarregada de criar as animações em Python. Em seguida, são gerados vários frames pertencentes à animação final, mas que não podem ser manipulados pela biblioteca. Os frames são gerados em formato PNG com dimensões de até 1920 x 1080 (full HD). Eles são posteriormente convertidos em um arquivo de vídeo codificado em H.264 com o uso do FFmpeg<sup>9</sup>. O processo completo utilizado na abordagem é mostrado na Figura 10.

---

<sup>9</sup> <http://www.ffmpeg.org/>



FIGURA 10 - PROCESSO DE CRIAÇÃO DAS NUVENS DE PALAVRAS ANIMADAS



Fonte: Elaborado pelo autor

O vídeo gerado ao fim do processo apresenta as características das pesquisas realizadas por um grupo ao longo dos anos. De forma simples e intuitiva, é possível observar, por exemplo, os temas mais pesquisados em um determinado ano, assim como as evoluções e consolidações dos temas pesquisados.

A técnica de Nuvem de Palavras utilizada comumente só apresenta uma espécie de fotografia dos dados em um momento determinado, mas não a dinâmica desses dados. É exatamente para cobrir essa lacuna que essa técnica é utilizada. Além disso, é possível aplicar a sincronização de vídeos relacionados a um mesmo período para evidenciar correlações entre os dados que poderiam estar escondidos quando apenas a visualização instantânea das Nuvens de Palavras é utilizada.

A fim de facilitar a visualização de informações associadas a várias nuvens de palavras animadas, foi desenvolvida uma aplicação Web em HTML5, para construir um mosaico da apresentação dos vídeos obtidos no processo da Figura 10. Esta estratégia permite revelar correlações temporais entre as informações apresentadas de forma simultânea na tela. Para isso, é necessário que todos os vídeos do mosaico estejam em sincronia. A aplicação HTML5 desenvolvida tem esse propósito e permite controlar a apresentação de todos os vídeos do mosaico por meio de uma única *timeline*. Como os vídeos foram gerados seguindo o mesmo processo, possuem a mesma duração (associada ao intervalo de tempo considerado

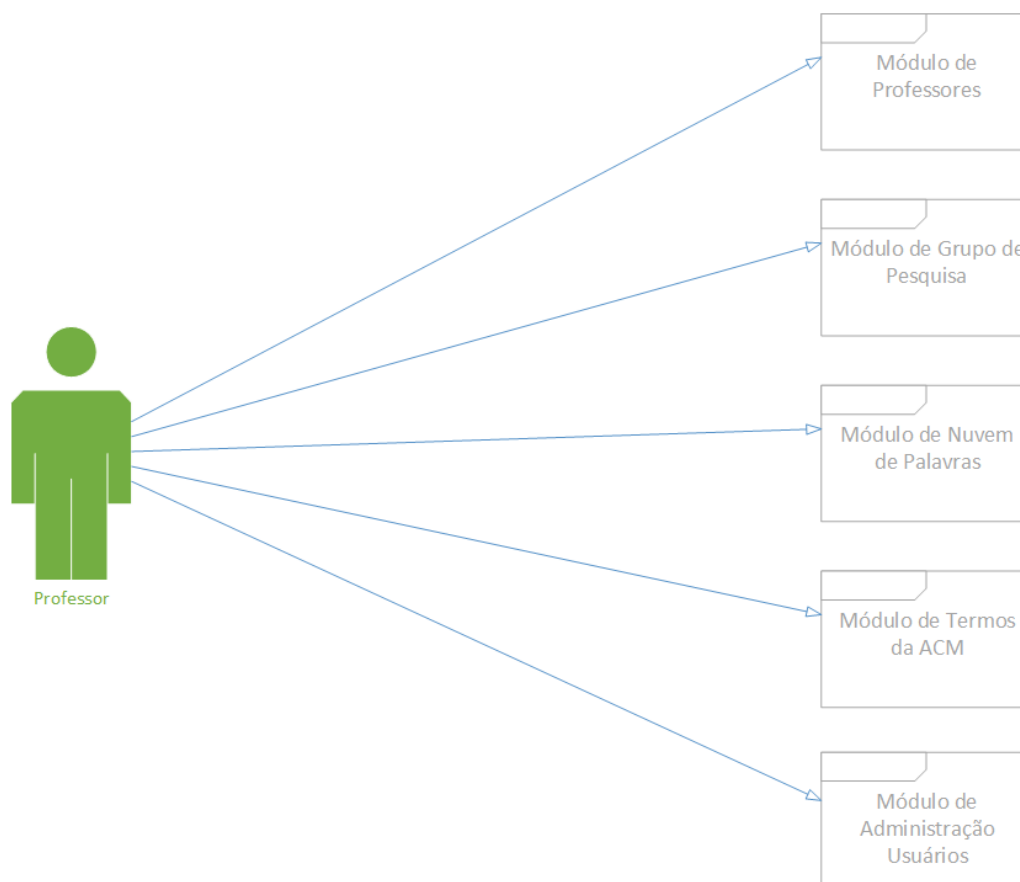
nos experimentos de 1995 a 2015) e com a mesma taxa e quantidade de frames, não foi necessário nenhum controle extra para que os vídeos mantenham uma sincronização perfeita após suas apresentações terem sido iniciadas ao mesmo tempo.

#### **4.6 Arquitetura do Sistema**

A representação da arquitetura do sistema fornece uma visão holística das abstrações que representam a estrutura e organização dos componentes do sistema (PRESSMAN; MAXIM, 2014).

A arquitetura do sistema ajuda a identificar os principais elementos estruturais, bem como descrever como o projeto está organizado. Nessa direção, a arquitetura deste sistema é formada pelo módulo de administração dos usuários, o módulo dos professores, o módulo dos grupos de pesquisa, o módulo das nuvens de palavras e o módulo dos termos da ACM, conforme o diagrama de pacotes da Figura 11.

FIGURA 11 - DIAGRAMA DE PACOTES VI-PROJECT

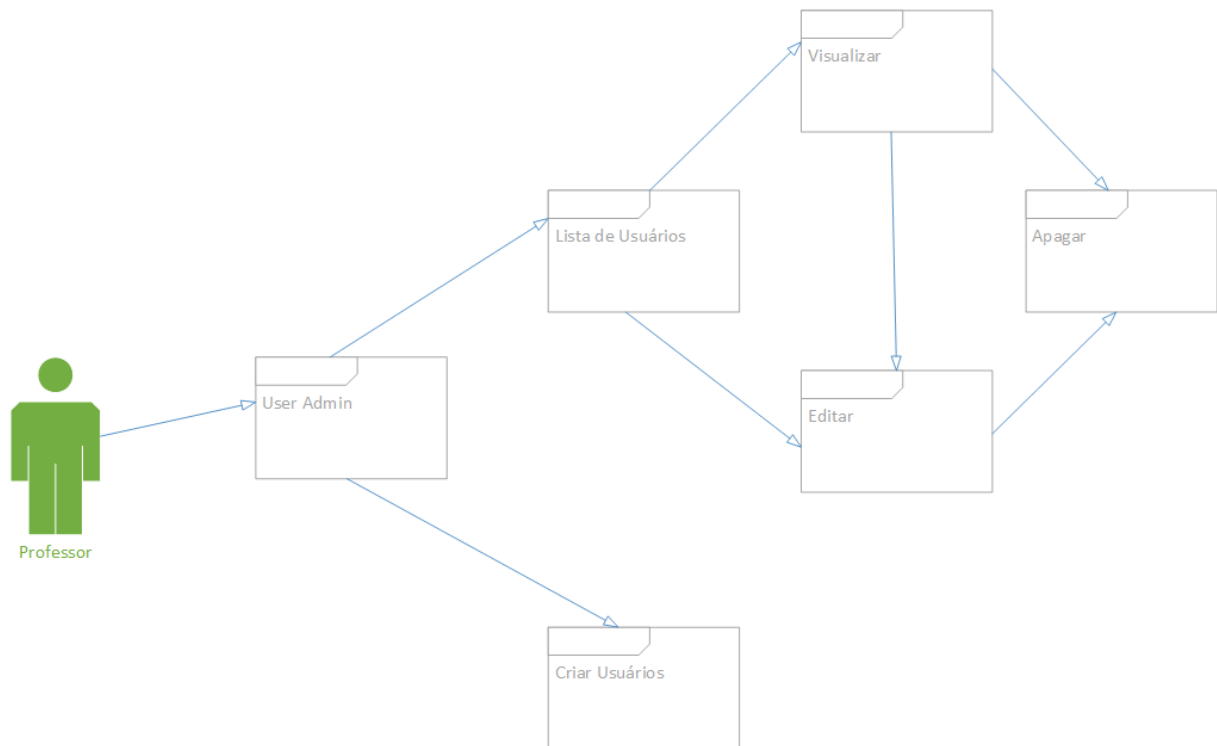


Fonte: Elaborado pelo autor

#### 4.6.1 Módulo de Administração dos Usuários

O Módulo de Administração dos Usuários foi criado baseado na biblioteca FOSUserBundle. Ela permitiu ter controle da segurança do sistema através do controle dos usuários e na administração dos perfis. O Módulo de Administração dos Usuários contém as funções básicas de criar, visualizar, editar e apagar usuários que terão acesso ao sistema VI-Project, tal como apresenta a Figura 12 no seu diagrama de pacotes.

FIGURA 12 - DIAGRAMA DE PACOTES MÓDULO ADMINISTRAÇÃO DOS USUÁRIOS

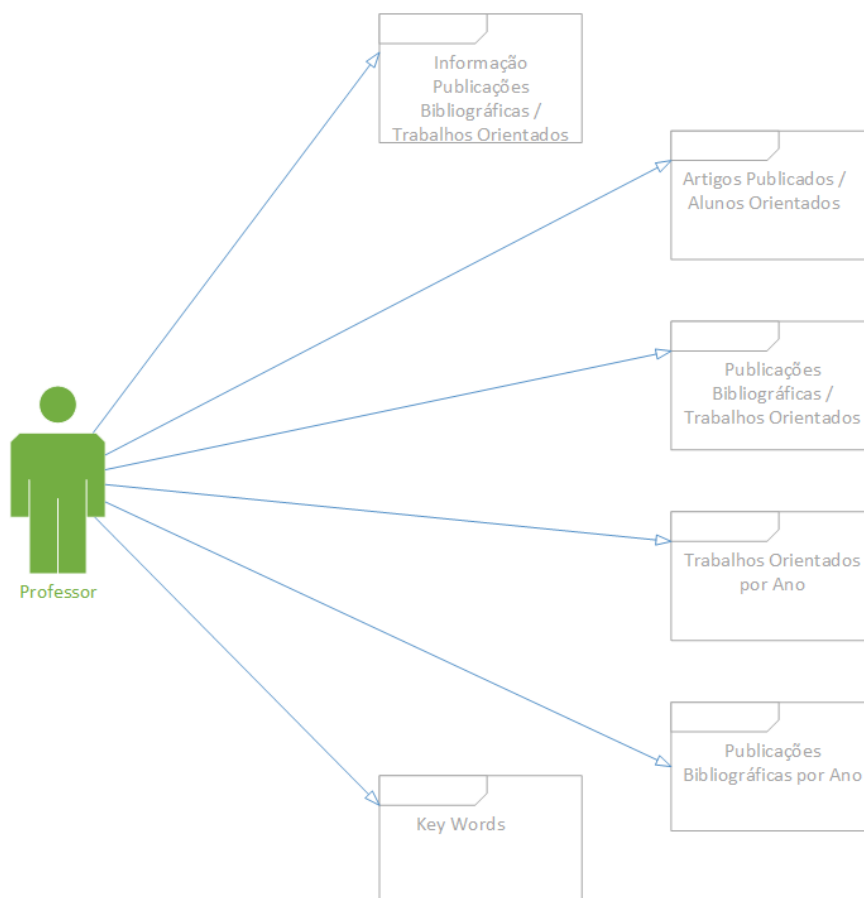


Fonte: Elaborado pelo autor

#### 4.6.2 Módulo dos Professores

No Módulo dos Professores é possível visualizar a informação individual das Produções Bibliográficas, assim como a informação dos Trabalhos Orientados de cada professor que pertence ao Programa de Pós-graduação em Informática. As visualizações de informações que se encontram nesse módulo são: Artigos Publicados / Alunos Orientados, Produções Bibliográficas / Trabalhos Orientados, Trabalhos Orientados, Produções Bibliográficas, Informação Produções Bibliográficas / Trabalhos Orientados, Key Words dos professores, conforme apresenta a Figura 13. Cada uma das visualizações que pertencem a este módulo foi criada com o processo da geração das Visualizações Estatísticas apresentado anteriormente.

FIGURA 13 - DIAGRAMA DE PACOTES MÓDULO PROFESSORES

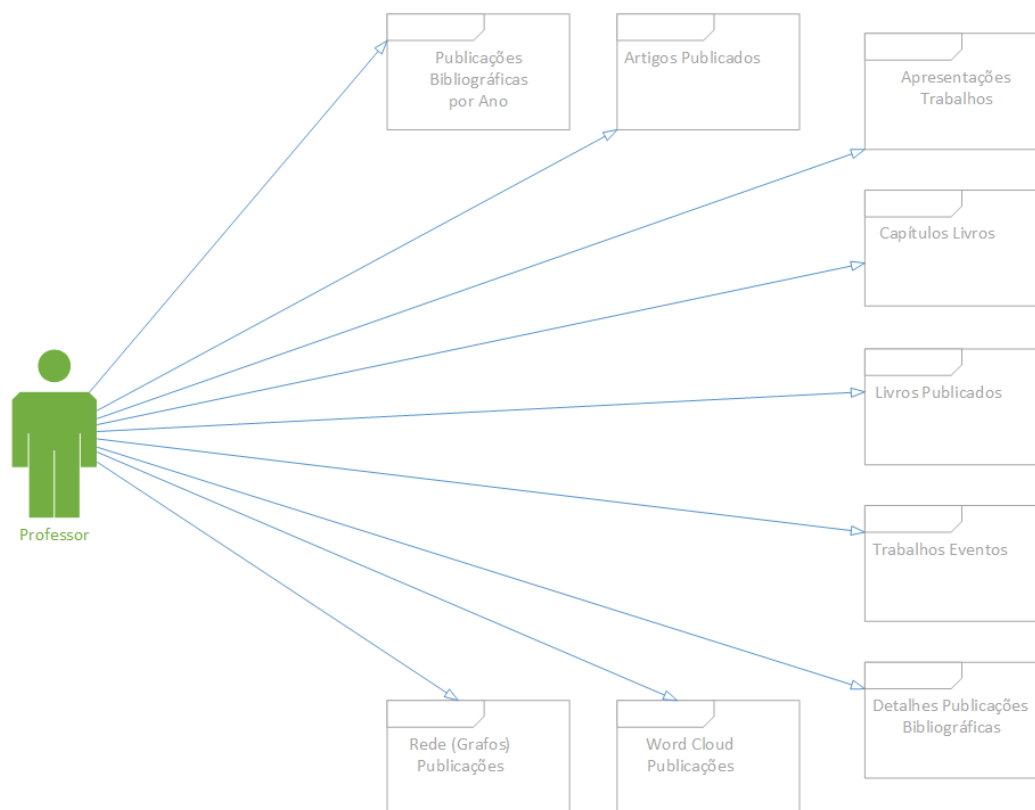


Fonte: Elaborado pelo autor

#### 4.6.3 Módulo dos Grupos de Pesquisa

O Módulo dos Grupos de Pesquisa faz referência às visualizações das informações das produções bibliográficas agrupadas dos professores que pertencem a grupos de pesquisa específicos. As visualizações de informações que se encontram nesse módulo são: Produções Bibliográficas por Ano, Artigos Periódicos, Apresentações Trabalhos, Capítulos Livros, Livros, Trabalhos Eventos, Detalhes Produções Bibliográficas, Word Cloud Publicações e Rede (Grafo) das Publicações, como apresenta a Figura 14. Cada uma das visualizações que pertencem a este Módulo foi criada com o processo da geração das visualizações estatísticas, apresentado anteriormente.

FIGURA 14 - DIAGRAMA DE PACOTES MÓDULO GRUPO DE PESQUISA

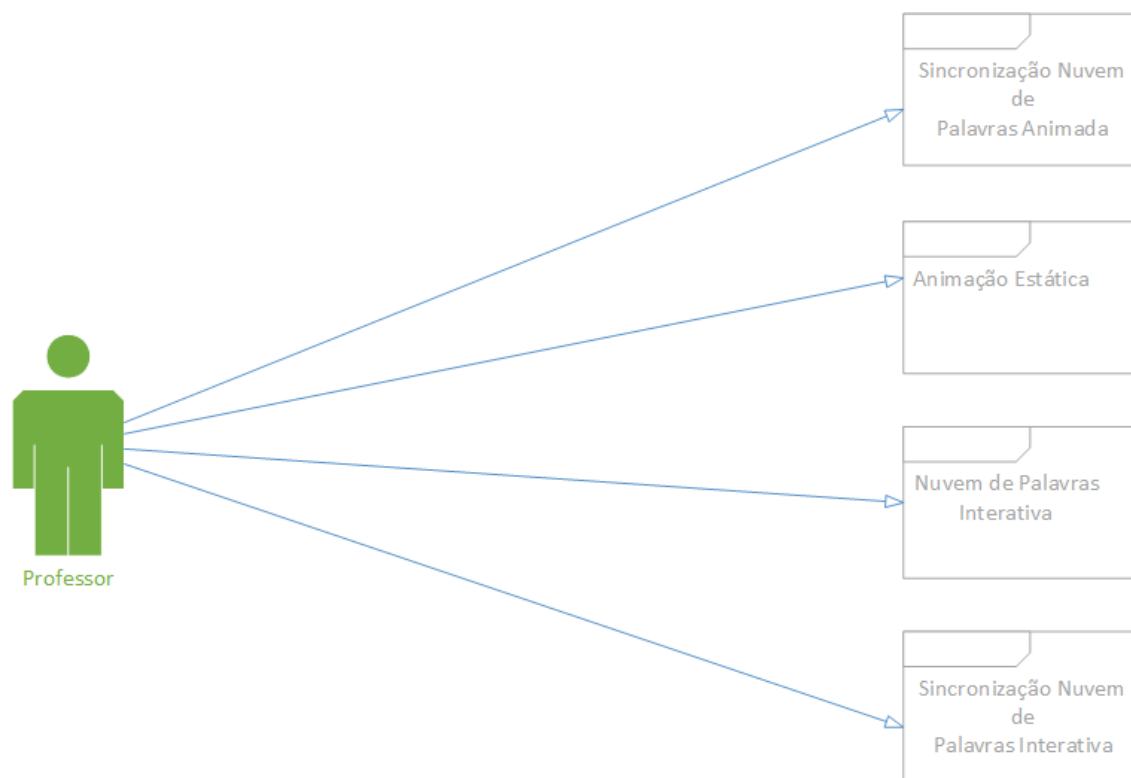


Fonte: Elaborado pelo autor

#### 4.6.4 Módulo das Nuvens de Palavras

O Módulo das Nuvens de Palavras contém diferentes técnicas de visualização de informação de Nuvens de Palavras para apresentar descrições visuais de uma coleção de dados textuais extraídos das palavras-chave utilizadas pelos professores em suas produções bibliográficas. As visualizações de informações que se encontram nesse módulo são: Sincronização de Nuvens de Palavras Animadas, Animação de Nuvem de Palavras Estáticas, Nuvem Interativa de Palavras e Sincronização das Nuvens Interativas de Palavras, como apresenta a Figura 15.

FIGURA 15 - DIAGRAMA DE PACOTES MÓDULO NUVENS DE PALAVRAS



Fonte: Elaborado pelo autor

A Sincronização de Nuvens de Palavras Animadas foi desenvolvida com uma aplicação Web em HTML5 para construir um mosaico da apresentação dos vídeos obtidos.

Na criação das Nuvens de Palavras Animadas são gerados frames que são utilizados para criar a Animação das Nuvens de Palavras Estática com uma aplicação Web em HTML5.

A Nuvem Interativa de Palavras e sua sincronização foram desenvolvidas com uma aplicação Web em HTML5 que ajudou a apresentar dinamicamente os temas relevantes pesquisados nos grupos e mostrar a informação desses temas através de um quadro ao clicar em alguma palavra. O processo de criação da Nuvem Interativa de Palavras foi apresentado anteriormente.

#### 4.6.5 Módulo dos termos da ACM

O Módulo dos Termos da ACM foi criado para observar o ingresso das palavras-chave de uma publicação, em quais categorias da ACM se localiza a publicação. Também serve para comprovar se o termo da palavra-chave é recomendado pela ACM e verificar se a categoria na que se encontra a palavra-chave é a correta ou esperada pelo pesquisador. Esta visualização de informação foi criada adaptando uma técnica de visualização de informação encontrada na revisão da literatura, especificamente a técnica de visualização de árvore.



## **5 EXPERIMENTOS E RESULTADOS**

## 5 EXPERIMENTOS E RESULTADOS

No capítulo anterior foi apresentado o VI-Project do ponto de vista da arquitetura do software, que permitiu detalhar o comportamento interno das principais visualizações e dos módulos que formam o sistema. Neste capítulo serão apresentados os resultados obtidos a partir das visualizações implementadas, além de algumas das contribuições do VI-Project.

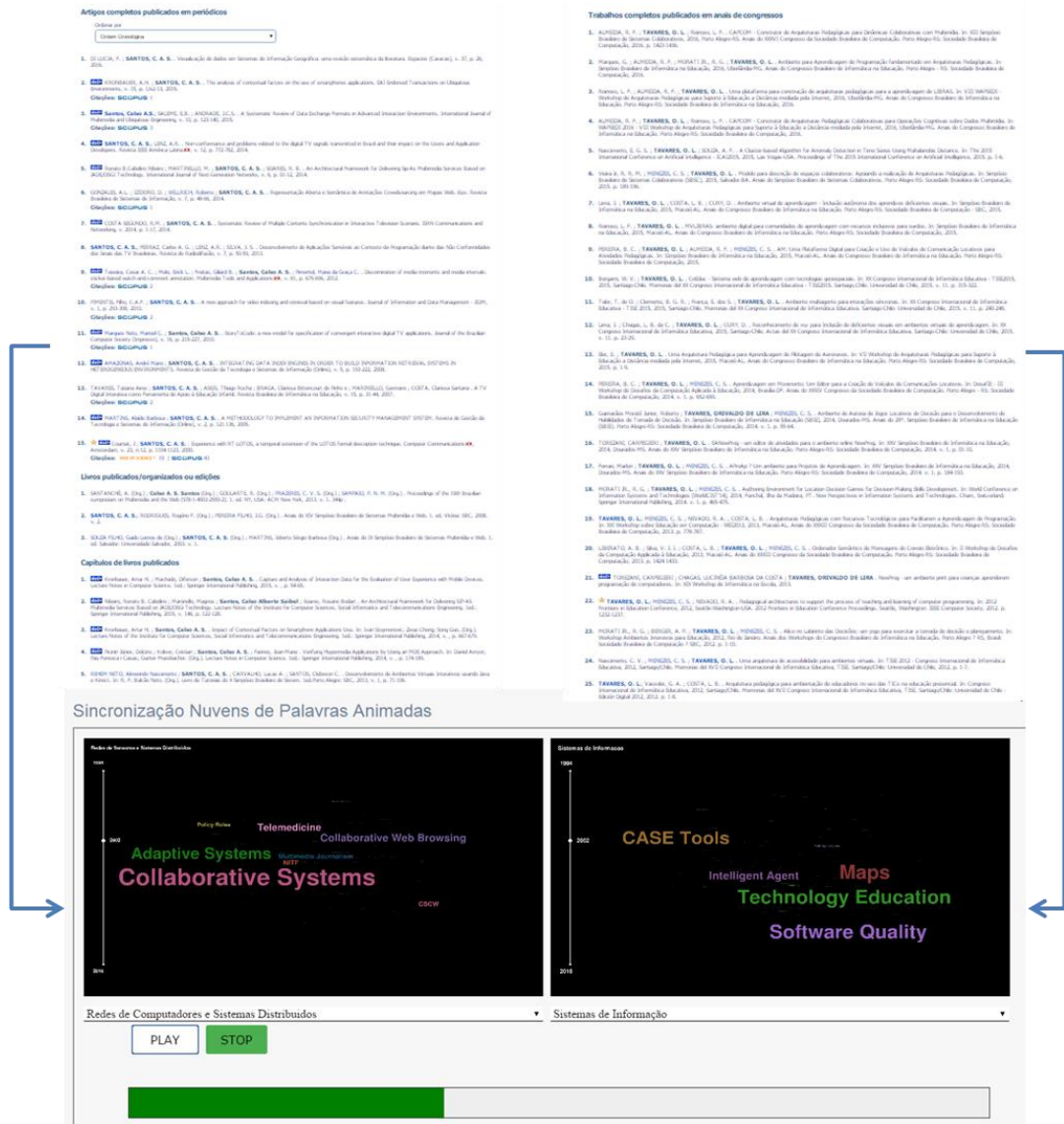
### 5.1 Resultados das Visualizações

Conforme exposto, os dados utilizados para gerar as visualizações foram obtidos diretamente dos CV-Lattes dos pesquisadores do grupo considerado. Nos CV, os dados são categorizados pelo tipo que representam (Produções Bibliográficas, Trabalhos Orientados, etc.) e correspondem a grandes quantidades de texto, a partir das quais não se pode facilmente extrair conhecimento. Eles podem ser vistos como um conjunto discreto de valores que não dizem nada e não são indicativos para a ação se não são processados e manipulados.

A partir da transformação desses dados, foi possível gerar informação com significados, relevância e valor, tornando-os mais úteis ao serem apresentados com o uso de diferentes técnicas de visualização. Conforme mostrado na Figura 16, essas técnicas são capazes de mudar a maneira como os usuários percebem os mesmos dados, permitindo a geração de conhecimento deles.

Do mesmo jeito, a transformação dos dados em informação e a apresentação da informação em técnicas de visualização permitem observar, com uma visão simples, uma informação agregada associada a uma grande quantidade de documentos.

FIGURA 16 - DADOS PARA A CRIAÇÃO DAS TÉCNICAS DE VISUALIZAÇÃO DE INFORMAÇÃO



Fonte: Elaborado pelo autor

### 5.1.1 Nuvens de Palavras

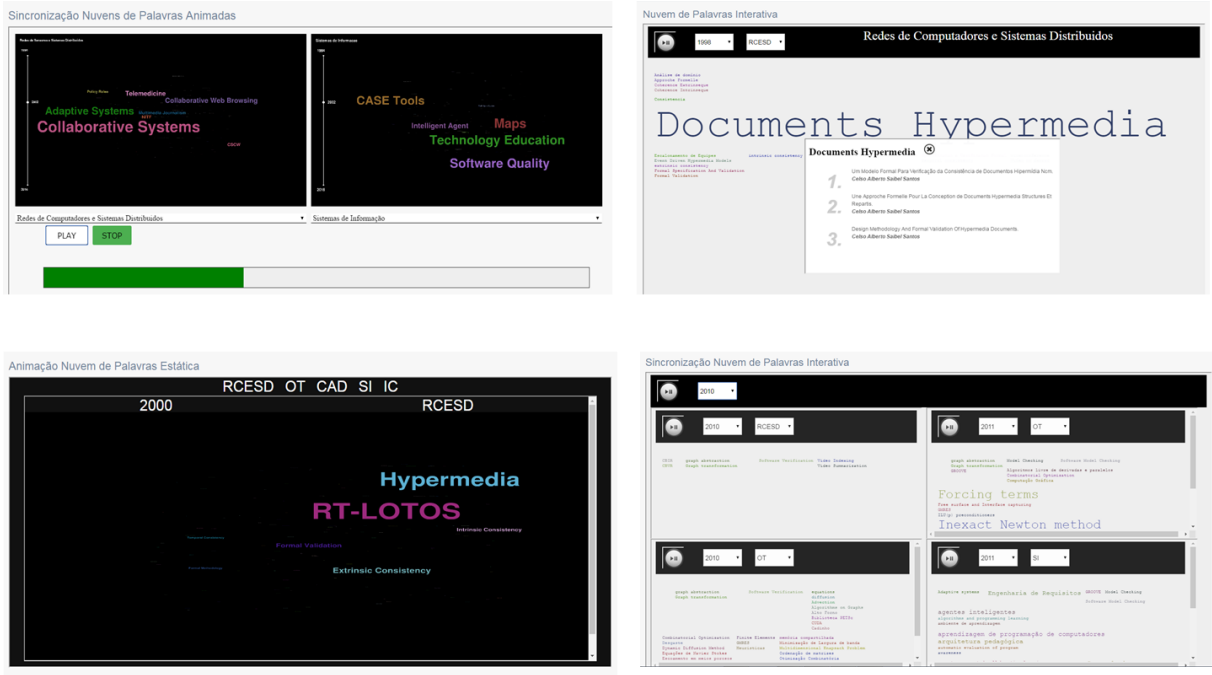
Nas diferentes Nuvens de Palavras apresentadas no site VI-Project (Figura 17) é possível observar quais os temas têm sido pesquisados nos últimos anos pelos Grupos de Pesquisa. Isso possibilita que se tenha uma visão clara de um assunto principal que centralizasse a atenção de todas as pesquisas do grupo ou a quantidade de publicações em um ano. Duas nuvens de palavras, em formato H.264 produzidas no contexto deste trabalho estão disponíveis em <http://bit.ly/2aS3DHy>.

O acompanhamento da evolução das palavras-chave associadas à produção acadêmica de um grupo pode evidenciar a consistência de uma determinada linha/área ao longo do tempo. Assim, quando uma ou mais palavras-chaves aparecem em destaque (tamanho maior da palavra) em determinados instantes da linha do tempo da Nuvem de Palavras, mas se tornam pequenos ou desaparecem em outros anos, existe um forte indício de que esse não foi um tema recorrente ou trata-se de um tema que atraiu a atenção momentânea dos pesquisadores durante certo período e depois perdeu o interesse por diferentes motivações (conclusão de um projeto, interesse comercial, fim de ciclos de desenvolvimento, entre outros). Além disso, quando o conjunto de palavras-chave relevantes relacionadas com um determinado grupo se mostra semelhante ao longo do tempo, isso pode indicar que a linha de pesquisa do grupo é consistente e bem definida.

A quantidade de palavras-chave exibidas com tamanho destacado (maior relevância) ao longo da linha do tempo também pode revelar pontos importantes sobre o perfil do grupo e de seus pesquisadores. Se as palavras-chaves relevantes são constantes, isso pode indicar que o grupo possui um perfil mais conservador, enquanto um número crescente de temas em evidência pode indicar que o grupo, além de dar continuidade aos temas já pesquisados, incorpora novos tópicos aos seus interesses de pesquisa ao longo do tempo.

Também, a sincronização das Nuvens de Palavras permite observar de uma vez só, como as informações de vários grupos de pesquisa evoluem no tempo e revelam correlações até então desconhecidas entre os dados.

FIGURA 17 - NUVENS DE PALAVRAS NO SITE VI-PROJECT



Fonte: Elaborado pelo autor

Ao comparar especificamente dois grupos de pesquisa – “Computação de Alto Desempenho” e “Inteligência Computacional” – pode-se descobrir os temas que têm sido pesquisados no mesmo período e quais são relevantes para ambos, como mostra a Figura 18. No ano 1998 o primeiro grupo teve destaque em três temas: *Performance Evaluation*, *Weightless Neural Networks* e *VLIW Architectures*. O segundo grupo destacou-se em *User Interfaces*, mas não com a mesma relevância que os temas do primeiro grupo.

FIGURA 18 - TEMAS PESQUISADOS NO ANO 1998



Fonte: Elaborado pelo autor

No ano 1999 observamos que o primeiro grupo continua com dois dos mesmos temas pesquisados anteriormente: *Performance Evaluation*, *Weightless Neural Networks*, mas aumentando sua relevância e é adicionado um novo tema de pesquisa: *DTSVLIW Architectures*. O segundo grupo nesse momento trabalhou com o tema *Active Design Documents*, sendo bem relevante, tal como indica o tamanho do termo na Figura 19.

FIGURA 19 - TEMAS PESQUISADOS NO ANO 1999



Fonte: Elaborado pelo autor

Para o ano 2002 os dois grupos mudaram completamente seu interesse de pesquisa, abordando o grupo 1 os temas *Computer Architecture* e *High-Performance Computing*, e o grupo 2 pesquisando diversos temas, mas o mais relevante foi *Design Rationale* como apresenta a Figura 20. É possível inferir com estas visualizações que o interesse nestes grupos de pesquisa não é constante no tempo e a mudança dos temas é considerável até esse momento, ou seja, os temas de pesquisa não estão completamente consolidados em ambos os grupos.

FIGURA 20 - TEMAS PESQUISADOS NO ANO 2002



Fonte: Elaborado pelo autor

Na Figura 21 observamos o ano 2006, o grupo 1 ainda continua com os temas pesquisados anteriormente: *Computer Architecture* e *High-Performance Computing*, mas o grupo 2 mudou consideravelmente seu interesse novamente, pesquisando nesse momento sobre: *Genetic Algorithms*, *Classification* (mais relevante) e *Automatic Classification of Documents*. Até o momento não é possível afirmar que o segundo grupo tem uma linha de pesquisa constante e bem definida, pois se observa a incorporação contínua de novos temas com o passar dos anos.

FIGURA 21 - TEMAS PESQUISADOS NO ANO 2006



Fonte: Elaborado pelo autor

Assim como apresenta a Figura 22, o segundo grupo no ano 2010 não teve temas relevantes de pesquisa e quase não publicou, mas o primeiro grupo mudou seus interesses pesquisando nesse ano: *Multi-label Text Categorization*, *VG-RAM WNN* (mais relevante), *Neural Networks* e *Portfolio Selection*, publicando uma grande quantidade de trabalhos.

FIGURA 22 - TEMAS PESQUISADOS NO ANO 2010



Fonte: Elaborado pelo autor



Para o ano 2015 (Figura 23), o primeiro grupo teve suas publicações baseadas em dois principais temas: *VG-RAM WNN* e *Robot Localization*. O segundo grupo desde o ano 2010 não teve grandes temas de pesquisa relevantes e pela quantidade de termos apresentados, podemos inferir que não teve muitas publicações nesse tempo.

FIGURA 23 - TEMAS PESQUISADOS NO ANO 2015



Fonte: Elaborado pelo autor

Assim como foi possível comparar os temas de pesquisa destes dois grupos, no *site*, é possível comparar e observar a evolução dos temas de pesquisa de outros grupos.

### 5.1.2 Rede (Grafos) das Publicações

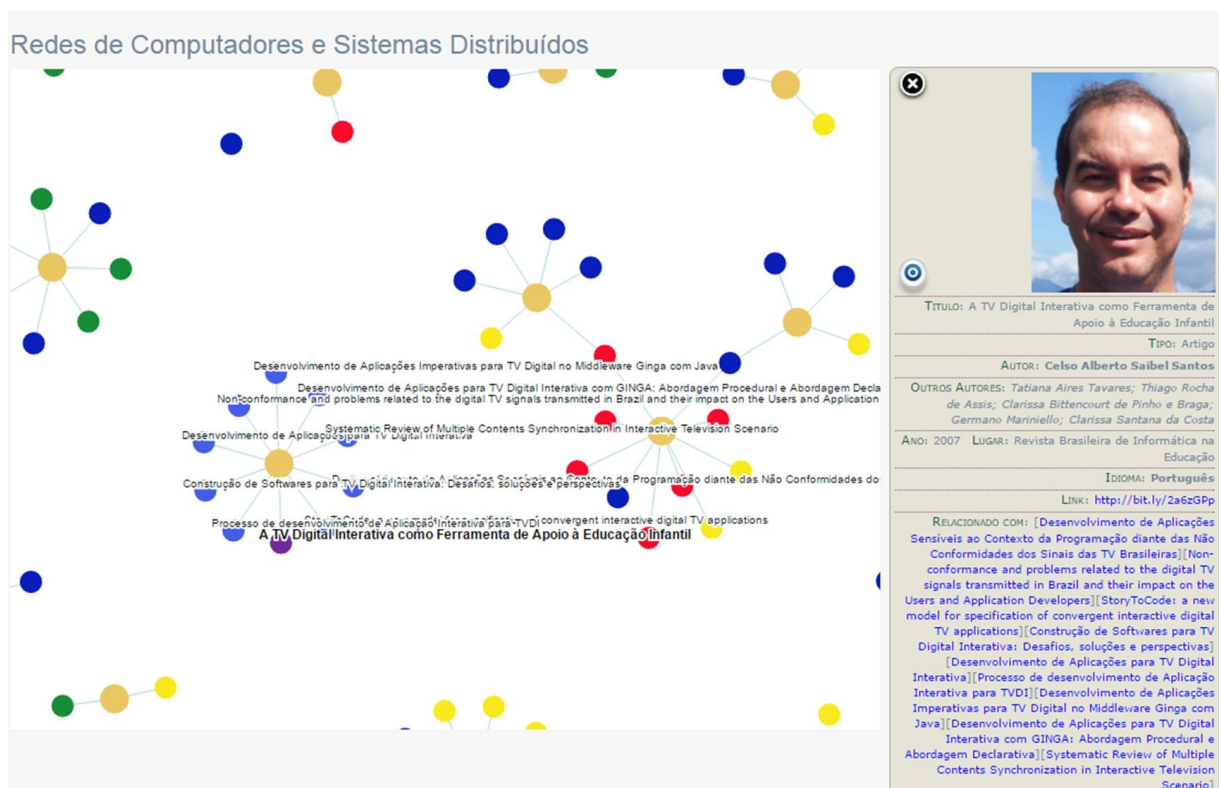
A visualização da Rede das Publicações permite separar as publicações em nós de acordo com sua classificação, sendo que a classificação é feita com o tema do que tratam as Produções Bibliográficas. Também é possível visualizar a relação entre os nós (Publicações) sem importar o tipo de publicação que ela é. Essa informação não é possível ser percebida com os CV Lattes dos professores de forma separada. Quando é selecionado algum nó, exibe-se um quadro com a informação detalhada dessa publicação, por exemplo: nome, autor, coautores, lugar da publicação, ano, link, trabalhos relacionados e outras informações.

Com as classificações principais é possível conhecer quais são as áreas de atuações dos Grupos de Pesquisa que se está visualizando, como apresenta a Figura 24.

Nessa visualização também é possível saber se um ou vários professores estão publicando na área de atuação do Grupo de Pesquisa ao que eles pertencem ou se estão trabalhando em novos temas. Também é possível observar aquelas publicações isoladas dos professores que não tem relação alguma com as áreas de atuações do Grupo de Pesquisa, se a quantidade de publicações isoladas de um professor é grande poderia ser possível que ele não pertença a esse grupo e se poderia sugerir mudar de Grupo de Pesquisa.

Cada cor apresentada representa um tipo de publicação bibliográfica. A verde representa as Apresentações dos Trabalhos; a azul representa os Artigos; a amarela representa os Capítulos de Livros e a vermelha, os Livros.

FIGURA 24 - REDE DAS PUBLICAÇÕES DE UM GRUPO DE PESQUISA



Fonte: Elaborado pelo autor

A descoberta obtida a partir dessa técnica de visualização foi que o grupo de pesquisa Computação de Alto Desempenho trabalha em temas diversos, mas que possui dois temas principais (*Neural Network* e *VG-RAM*), associados à maioria de suas publicações mais recentes. A maior parte dos artigos de periódicos está relacionada ao tema *SUPG*. Além disso, existe uma grande quantidade de publicações bibliográficas isoladas, sem relação direta com os temas já consolidados no grupo.

Para o grupo de pesquisa Inteligência Computacional, foi descoberto com essa técnica que seu principal tema de pesquisa é *Ontology*. Por outro lado, o grupo tem exibido certa consolidação em temas como *Learning*, *Facility Layout Problem*, *Neural Network*, *VG-RAM* e *Particle Swarm*. Uma descoberta importante foi que alguns professores que pertencem ao grupo de pesquisa Computação de Alto Desempenho também pertencem ao grupo Inteligência Computacional, por isso, alguns dos temas listados aparecem em ambos os grupos.

Também foi possível descobrir que no grupo de pesquisa Otimização não se observa um tema de pesquisa ressaltante. Os temas de pesquisa *Finite Element*, *Facility Layout Problem*, *SUPG* e *Graph* têm praticamente a mesma relevância e são os mais destacados. O grupo de pesquisa possui alguns professores que também atuam nos grupos de pesquisa Computação de Alto Desempenho e Inteligência Computacional. Além disso, é possível verificar que o grupo ainda não possui um tema de pesquisa consolidado até o momento.

No caso do grupo de pesquisa Redes de Computadores e Sistemas Distribuídos foi possível descobrir que o tema de pesquisa consolidado é *Multimedia*. Mas, também tem outros temas de pesquisa que são importantes como *TV*, *Context-Aware* e *Network*. Esse grupo tem poucas publicações isoladas, então é possível dizer que sua produção bibliográfica é consistente no tempo. Foi descoberto um crescimento nas publicações associadas aos temas de pesquisa *Map*, *Graph* e *Smartphone*. Além disso, os professores desse grupo não pertencem simultaneamente a múltiplos grupos de pesquisa.

Para finalizar, no grupo de pesquisa Sistemas de Informação foi descoberto que o tema fortemente consolidado, assim como o de Inteligência Computacional, também é *Ontology*. A maior parte das publicações está associada a esse tema. Foi possível observar que outros temas majoritariamente pesquisados são *Learning* e *Knowledge*. Nesse grupo não existem muitas publicações isoladas, sendo que a maioria pertence a um tema de pesquisa específico.

### 5.1.3 Visualizações Estatísticas

Para conhecer, com detalhes, tanto os Artigos em Periódicos, as Apresentações dos Trabalhos, os Capítulos dos Livros, os Livros e os Trabalhos em Eventos, foram criadas as seguintes técnicas: um quadro com cálculos, um gráfico circular e um gráfico de barras (estudados no capítulo 2) que vão gerar depois de ser selecionado o Grupo de Pesquisa do que se quer visualizar a informação.

No quadro com os cálculos é apresentado o Total dos Trabalhos Publicados do grupo de pesquisa selecionado, o Número de Professores do Grupo, a Média dos Trabalhos Publicados de cada Professor e a Média dos Trabalhos do Ano avaliado. No gráfico circular é possível visualizar as quantidades totais dos Trabalhos Publicados de cada professor do Grupo de Pesquisa. E no gráfico de barras, é possível evidenciar a quantidade de Trabalhos Publicados de cada professor em cada ano, tal como é apresentado na Figura 25.

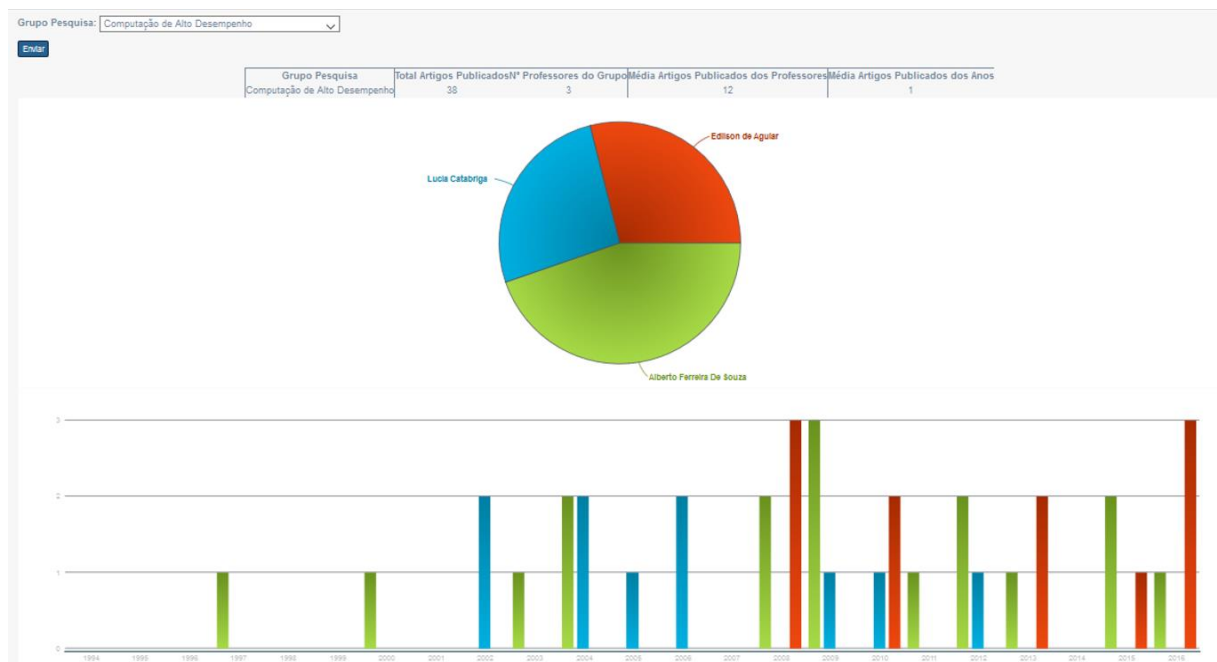
Na visualização dos Detalhes das Produções Bibliográficas é possível mostrar um quadro com o cálculo do Total das Produções Bibliográficas, o cálculo do Nº de Professores do Grupo que tem Produções Bibliográficas e o cálculo da Média das Produções Bibliográficas dos Professores em um ano específico. Além disso, apresenta um gráfico circular com o total das quantidades das produções bibliográficas de cada professor. Finalmente, é apresentado um gráfico de barras onde se observa a quantidade total das Produções Bibliográficas, classificadas pelo tipo de publicação de cada professor.

Essas visualizações ajudam a entender qual é o professor de um Grupo de Pesquisa que tem maior quantidade de trabalhos publicados, também qual é o tipo

de trabalho publicado majoritariamente pelo grupo, indicando uma potencial liderança no grupo analisado.

No exemplo de visualização do grupo de pesquisa Computação de Alto Desempenho apresentado na figura a seguir, consideram-se três professores. A cada um deles foi associada uma cor que representa as suas produções. Com uma visão simples se pode descobrir que o professor que tem a cor verde associada é quem tem maior quantidade de publicações, tal como apresenta o gráfico circular. Também no gráfico de barras se observa que o professor da cor verde tem maior quantidade de publicações porque tem publicado em mais anos. Por outro lado, o professor da cor vermelha tem produção recente mais elevada. E, em uma pesquisa posterior, descobriu-se que o professor da cor vermelha passou a integrar o Programa de Pós-graduação e o grupo de pesquisa em questão após 2008, ano a partir do qual suas publicações passaram a ser contabilizadas para a análise.

FIGURA 25 - VISUALIZAÇÃO ESTATÍSTICA DE UM GRUPO DE PESQUISA

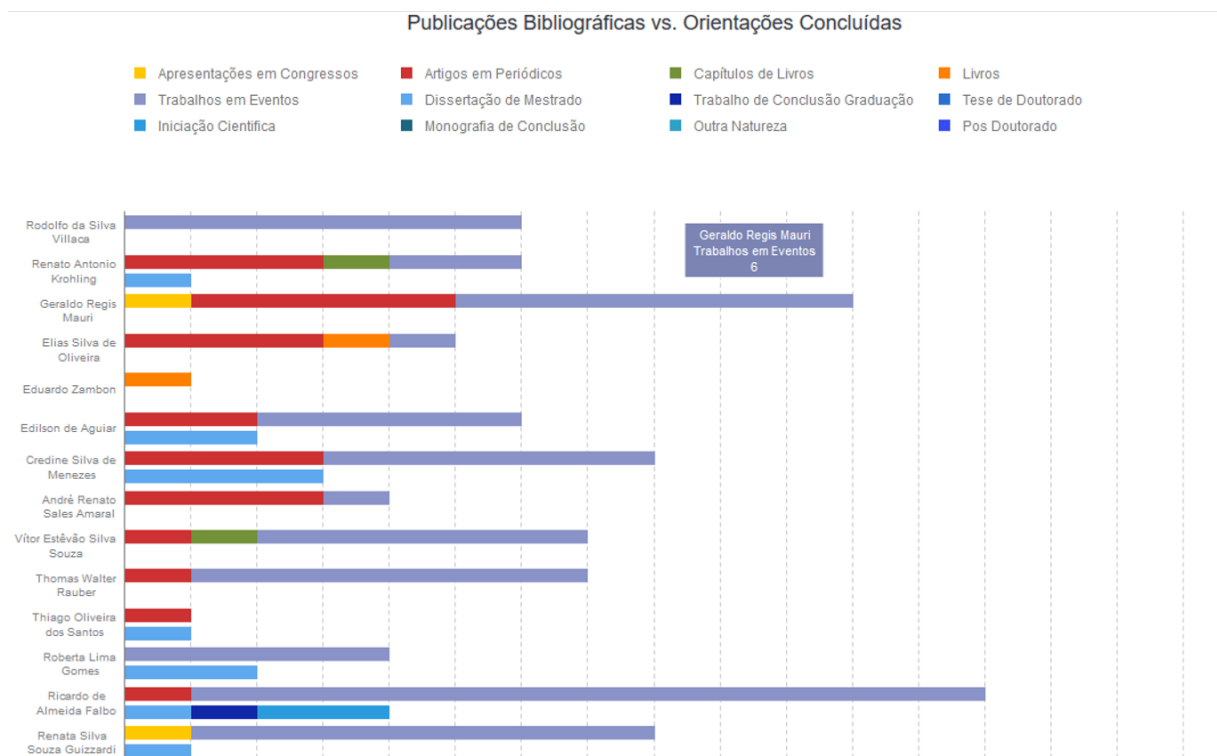


Fonte: Elaborado pelo autor

Outras visualizações estatísticas fazem a comparação da quantidade das produções bibliográficas com a quantidade de trabalhos orientados que os professores tiveram em um ano específico através de um gráfico de barras horizontal. Também é possível observar a quantidade específica de cada tipo de

publicação tanto dos professores individualmente, quanto dos grupos de pesquisa, assim como a quantidade específica de cada tipo de trabalho orientado. Contrastar essas quantidades pode evidenciar a atividade que um professor dedica seu tempo, se um professor tem maior quantidade de trabalhos orientados e menor quantidade de produções bibliográficas. O resultado dessa comparação pode indicar situações indesejadas para produção do programa. De um lado, se os temas de pesquisa orientados não têm gerado publicações relevantes como esperado, isso pode indicar que existem problemas relacionados aos temas escolhidos, à forma de orientação ou à dedicação de pesquisadores e alunos. Do outro lado, se um professor publica muitos trabalhos relevantes, mas orienta poucos alunos, isso pode indicar problemas na seleção de alunos para o tema de interesse ou menor interesse e dedicação a este tipo de atividade. O gráfico de barras para classificar tanto do tipo de trabalho orientado quanto do tipo de publicação bibliográfica também é útil para explicitar os interesses de pesquisa do professor. A Figura 26 ilustra a comparação das publicações bibliográficas e dos trabalhos orientados dos professores no ano 2013 usando esta técnica de visualização.

FIGURA 26 - VISUALIZAÇÃO PRODUÇÕES BIBLIOGRÁFICAS / TRABALHOS ORIENTADOS 2013



Ao se observarem as publicações bibliográficas e as orientações concluídas pelos professores do PPGI desde o ano 2010, foram descobertas evidências dos interesses no trabalho dos professores.

No ano de 2010 (Figura 27) foi evidente que os professores G. R. Mauri (15 trabalhos publicados *versus* 0 orientações concluídas) e Giancarlo Guizzardi (25 trabalhos publicados *versus* 2 orientações concluídas) exibem uma tendência de dirigir seus esforços para a publicação de trabalhos e não para a orientação de estudantes. Observa-se também que a tendência dos dois professores foi de publicar trabalhos em eventos, característica marcante dos programas ligados à área de computação. Por outro lado, nos professores O. L. Tavares (2 publicações *versus* 9 orientações concluídas) e M. Martinello (2 publicações bibliográficas *versus* 7 orientações concluídas) fica evidente a dedicação maior às atividades de orientação do que à produção de trabalhos científicos. Professores como C. A. S. Santos (8 publicações bibliográficas *versus* 7 orientações concluídas) e M. C. Rangel (3 publicações bibliográficas *versus* 3 orientações concluídas) mantêm equilibrado seu trabalho nessas duas responsabilidades.

No ano de 2011 (Figura 28) foi descoberto que os professores G. R. Mauri e Giancarlo Guizzardi são os que destacam novamente no grupo estudado, por terem muitas publicações bibliográficas e poucas orientações concluídas, mantendo a tendência observada anteriormente. O docente M. Martinello manteve a tendência evidenciada no ano anterior, indicando maior sucesso nas atividades de orientação de alunos, sendo no ano de 2011 o único professor com mais orientações concluídas do que trabalhos publicados.

FIGURA 27 - VISUALIZAÇÃO PRODUÇÕES BIBLIOGRÁFICAS / TRABALHOS ORIENTADOS 2010

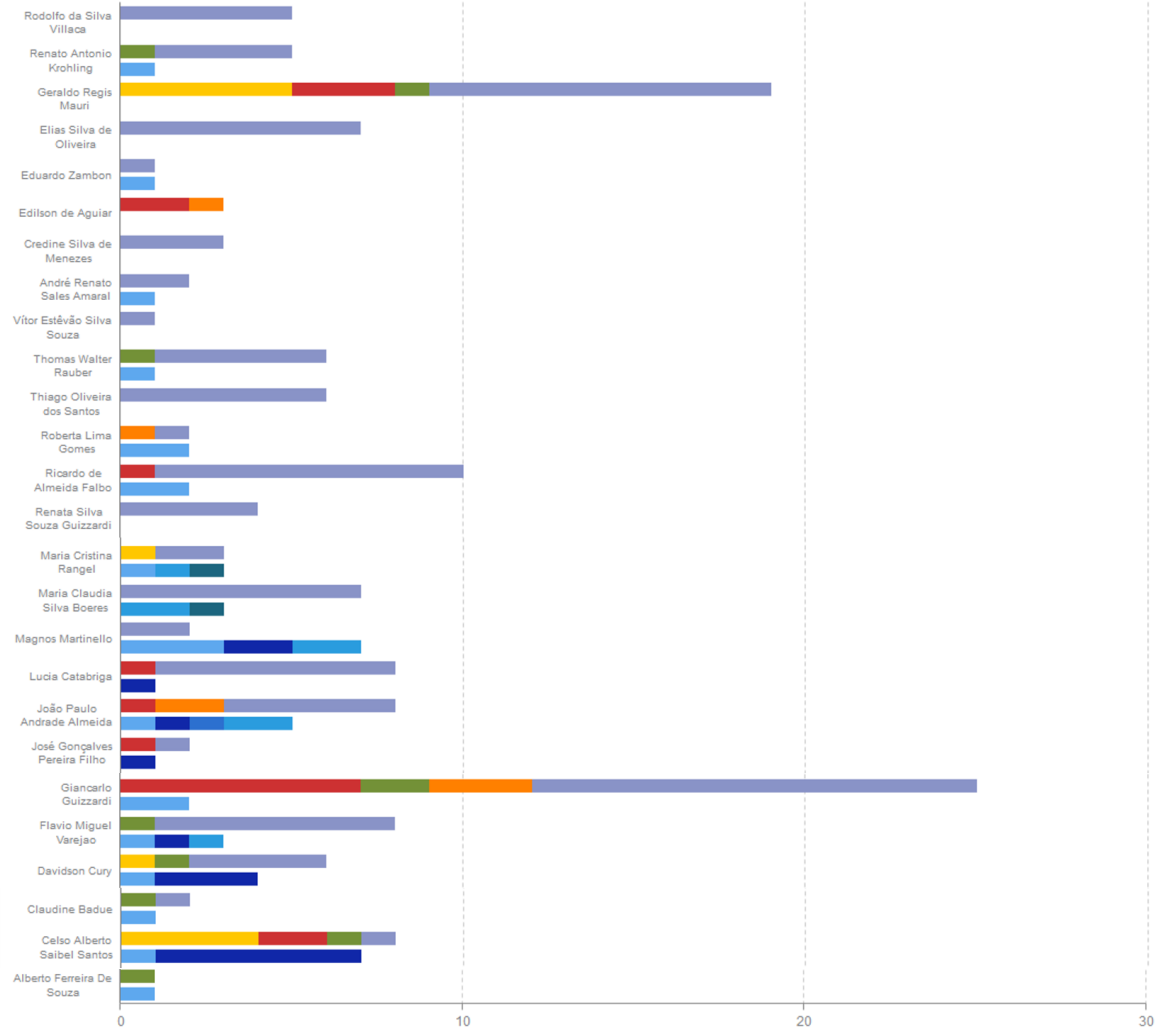
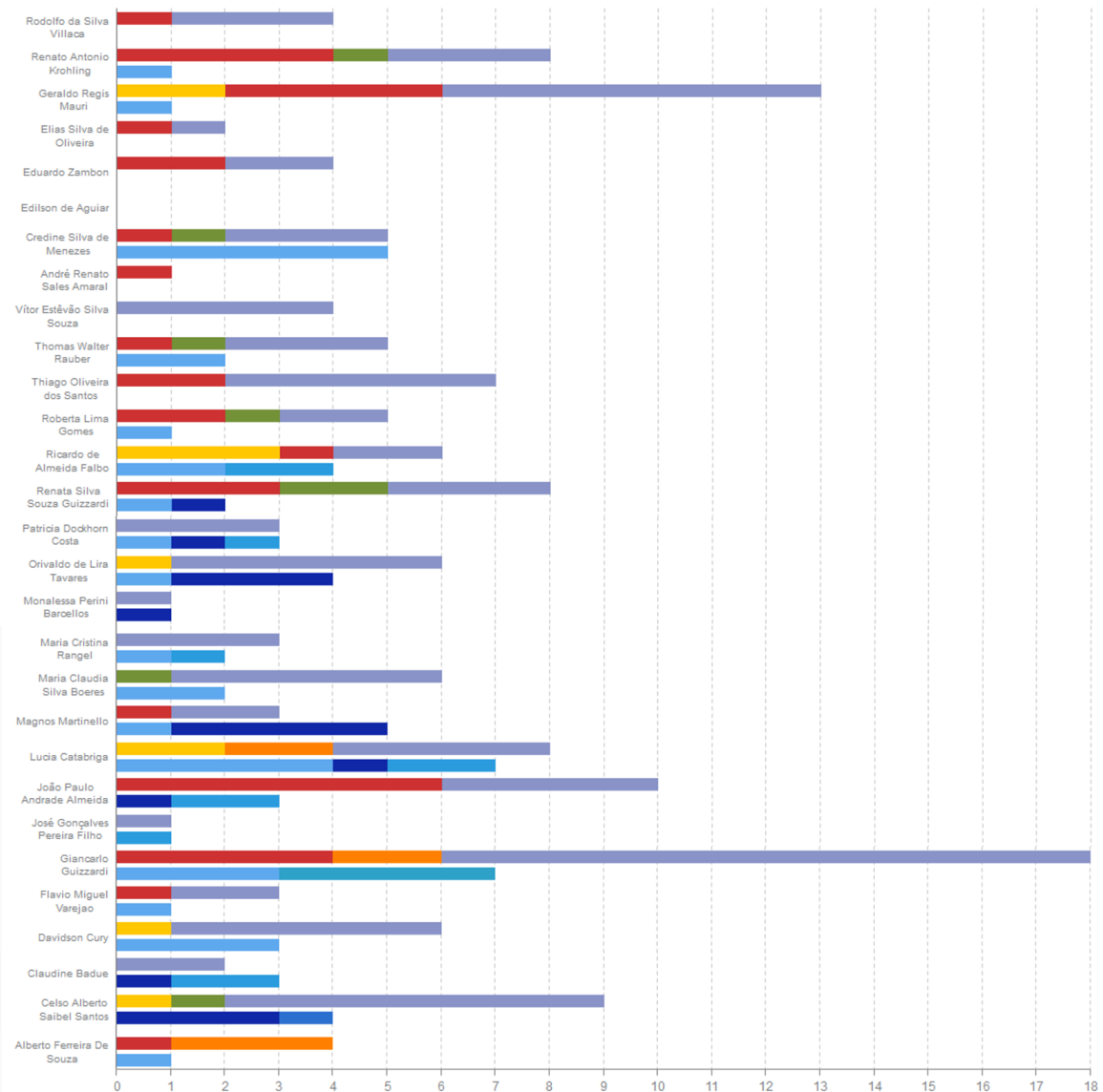


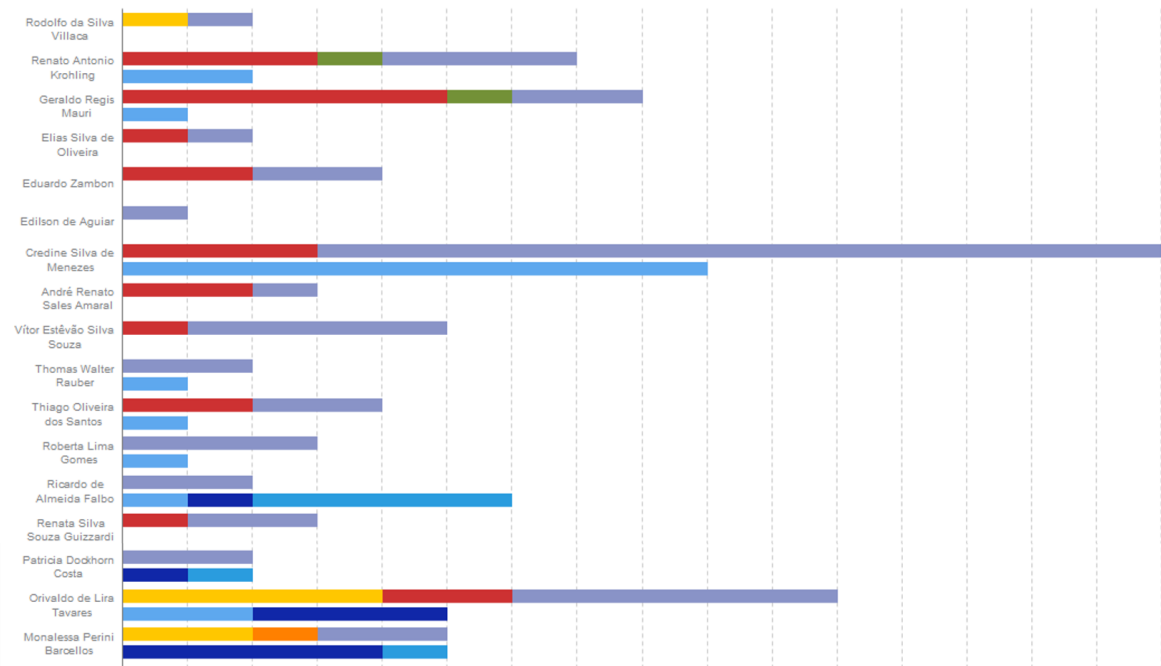


FIGURA 28 - VISUALIZAÇÃO PRODUÇÕES BIBLIOGRÁFICAS / TRABALHOS ORIENTADOS 2011



O ano de 2012 (Figura 29), sendo o terceiro ano consecutivo analisado e a com as mesmas evidências de 2010 e 2011, é possível afirmar que os professores G. R. Mauri e G. Guizzardi orientam poucos alunos e têm foco de atuação nas publicações. Por outro lado, alguns professores exibem tendências não observadas nos anos anteriores do triênio, tais como o professor C. S. Menezes, que possui muitas publicações bibliográficas e também muitas orientações concluídas nesse ano.

FIGURA 29 - VISUALIZAÇÃO PRODUÇÕES BIBLIOGRÁFICAS / TRABALHOS ORIENTADOS 2012

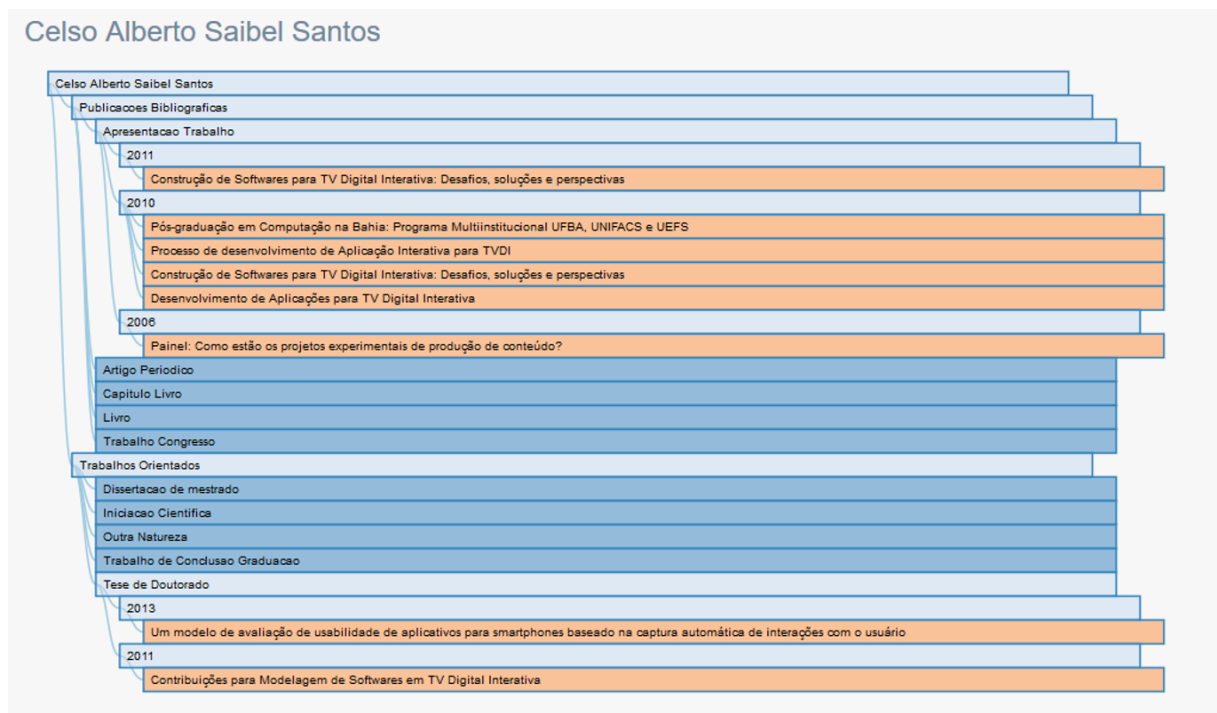


Ao prosseguir com o estudo correlacionando as atividades de publicações bibliográficas e as orientações concluídas até 2016, encontramos as mesmas tendências apresentadas anteriormente, no triênio 2010-2011-2012. Foi descoberto também que os números apresentados dos trabalhos orientados são especificamente das orientações concluídas nesse ano, isso não quer dizer que os professores não trabalharam com alunos, mas houve uma necessidade de maior tempo para conclusão dos estudos. Por exemplo, uma orientação de mestrado requer pelo menos dois anos para culminar e uma orientação de doutorado requer quatro anos. Por isso, a tendência sempre será de ter maior quantidade de publicações bibliográficas. Outra tendência descoberta foi que os professores majoritariamente publicam seus trabalhos em eventos ou congressos.

Concluindo, a forma de visualização de gráfico de barras é uma ferramenta adequada para apresentar comparativamente o desempenho de todos os professores do PPGI-UFES, permitindo evidenciar a correlação entre orientações e trabalhos publicados. Deve-se notar, porém, que o desempenho observado de alguns professores foi afetado na análise pelo seu recente ingresso no programa. Em consequência, não foi possível nem extrair qualquer tipo de tendência das suas produções adequadamente no período.

Se existisse o interesse de conhecer o nome tanto das produções bibliográficas, quanto dos trabalhos orientados que os professores tiveram individualmente de cada ano, eles podem ser apresentados através da visualização conjunta Informação Produções Bibliográficas / Trabalhos Orientados. Essa forma de visualização é apresentada na Figura 30.

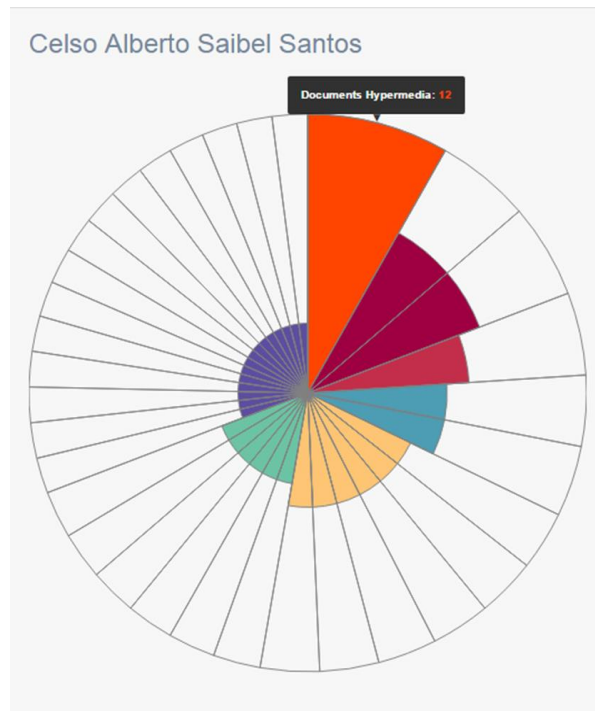
FIGURA 30 - VISUALIZAÇÃO INFORMAÇÃO PRODUÇÕES BIBLIOGRÁFICAS / TRABALHOS ORIENTADOS



Fonte: Elaborado pelo autor

Outra forma interessante para visualizar os temas pesquisados pelos professores é através de um gráfico circular. Ele apresenta gráfica e dinamicamente os principais temas que os professores têm pesquisado ao longo do tempo. O gráfico vai mostrar partes maiores que outras, enquanto as partes forem maiores significa que as produções bibliográficas principalmente têm relação com esse termo. Assim, o gráfico vai representar até o nível menor dos temas pesquisados pelos professores, como é apresentado na Figura 31.

FIGURA 31 - VISUALIZAÇÃO KEYWORDS DOS PROFESSORES

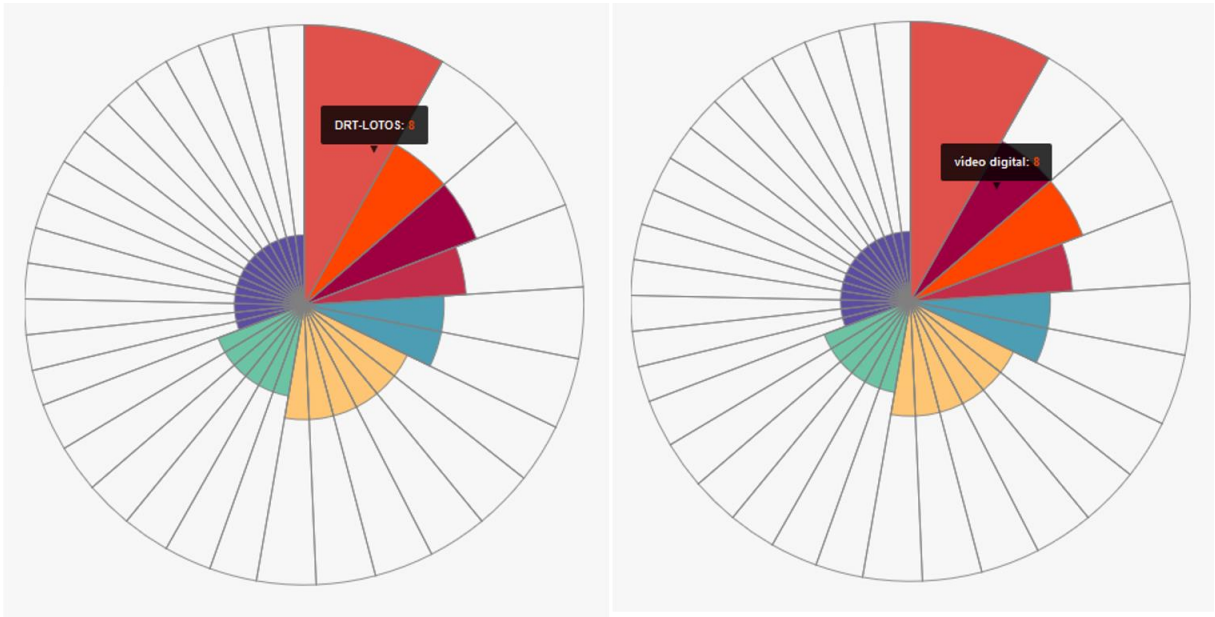


Fonte: Elaborado pelo autor

No caso do professor C. A. S. Santos, é possível observar que seu *Keyword* utilizado majoritariamente é *Documents Hypermedia*. Seguido pelos *Keywords RT-LOTOS* e *Digital Video*, como apresenta a Figura 32. Se continuarmos percorrendo o gráfico circular, será possível conhecer o resto das palavras-chave que foram utilizadas pelo professor.

Foi descoberto que os *Keywords* utilizados pelos professores têm íntima relação com os temas pesquisados. É possível dizer que os *Keywords* são também os temas de interesse deles.

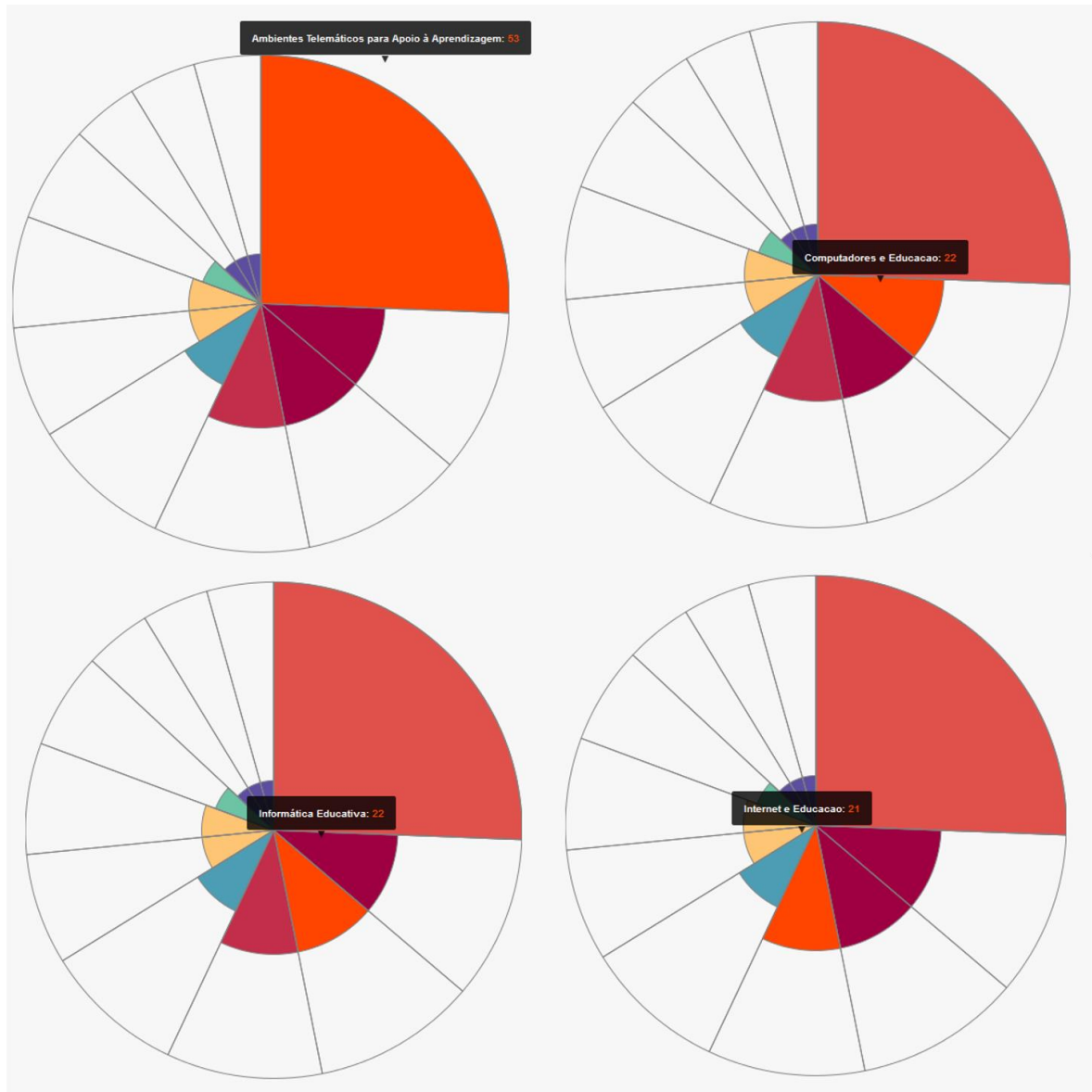
FIGURA 32 - VISUALIZAÇÃO KEYWORDS DOS PROFESSORES



Fonte: Elaborado pelo autor

No caso do professor C. S. Menezes foi descoberto que seus temas de pesquisa estão relacionados principalmente com Informática na Educação, conforme apresentado na Figura 33. Seu tema principal pesquisado e bastante relevante em comparação com os outros é Ambientes Telemáticos para Apoio à Aprendizagem, como é possível observar. Logo, ele possui três temas com a mesma importância e entre os primeiros temas pesquisados, eles são Computadores e Educação, Informática Educativa e Internet e Educação. Se continuarmos analisando o gráfico circular, será possível conhecer o resto dos temas pesquisados pelo professor.

FIGURA 33 - VISUALIZAÇÃO KEYWORDS DOS PROFESSORES



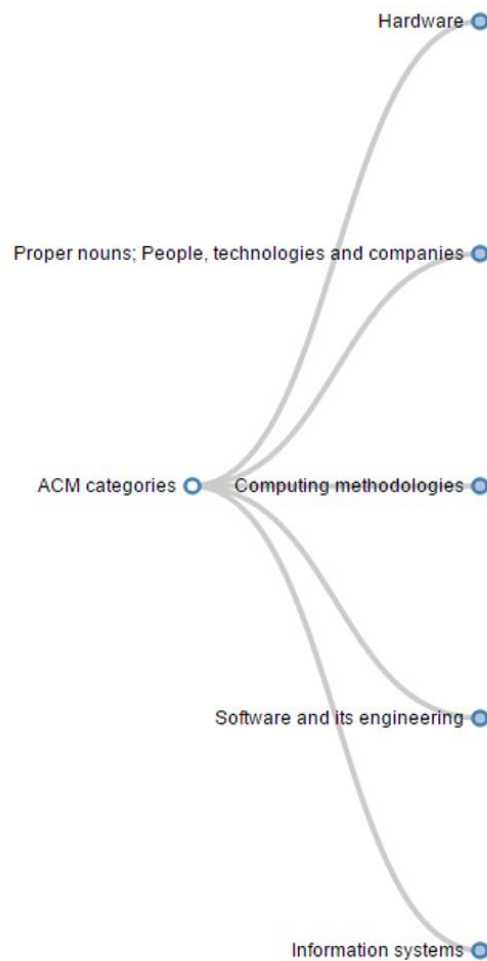
Fonte: Elaborado pelo autor

#### 5.1.4 Visualização dos Termos da ACM

Nesta visualização é possível observar em qual (ou quais) categoria(s) se insere um trabalho de pesquisa mediante as palavras-chave. Também, permite ver se o termo utilizado como palavra-chave é um termo aceito pela ACM, além de verificar se as categorias são as esperadas pelos pesquisadores. Essa visualização ajuda muito ao professor saber se o seu trabalho está categorizado corretamente.

No caso da palavra-chave *video*, foi descoberto que é uma palavra válida pela ACM, que pertence a diversas categorias. Essas categorias são: *Hardware*, *Proper nouns; People, technologies and companies*, *Computing methodologies*, *Software and its engineering*, *Information systems*, tal como apresenta a Figura 34.

FIGURA 34 - VISUALIZAÇÃO TERMOS DA ACM

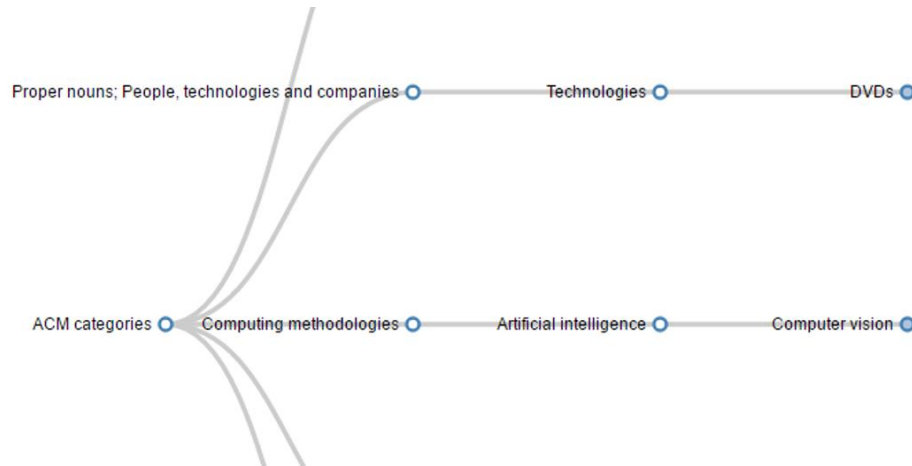


Fonte: Elaborado pelo autor

Também é possível conhecer com detalhes as subcategorias que pertence a palavra-chave ingressada. Neste caso, ao expandir as categorias principais, observa-se que a palavra *video* é sugerida em diferentes subcategorias. Algumas delas são apresentadas na Figura 35.

Esta visualização também apoia a selecionar de forma correta as palavras-chave do trabalho de pesquisa, sendo categorizadas e sugeridas por uma organização reconhecida internacionalmente.

FIGURA 35 - VISUALIZAÇÃO TERMOS DA ACM



Fonte: Elaborado pelo autor

## 5.2 Considerações do capítulo

O uso de técnicas adequadas de visualização de Informação permite o melhor entendimento do que se quer visualizar e também permite a extração de conhecimento dessas informações. Com o estudo, é possível afirmar que o título e as palavras-chave de Produções Bibliográficas, sendo bem analisados e apresentados, podem dizer mais dos simples campos de texto sem sentido.

Uma discussão detalhada sobre a escolha e uso das técnicas implementadas nesse trabalho foi feita ao longo do capítulo. As técnicas foram utilizadas na análise das informações oriundas do CV Lattes do conjunto de pesquisadores do PPGI-UFES, permitindo a extração de conhecimento (em termos de tendências, evidências, entre outras formas) sobre estes pesquisadores e também sobre o programa.

O trabalho não pretende substituir as formas de visualização já disponíveis para os CV Lattes dos pesquisadores. Ele pretende complementar as informações, apresentando evidências que não são possíveis de serem observadas individualmente.



## **6 EXPERIMENTOS E RESULTADOS FUTUROS**

## 6 CONCLUSÃO E TRABALHOS FUTUROS

A visualização de informação permite descrever, explorar e analisar a informação, com o objetivo de descobrir e ampliar o conhecimento. Este trabalho tratou da questão da extração de conhecimento a partir da visualização de informações.

As técnicas de visualização de informação são desenvolvidas a partir de dados qualitativos e quantitativos. Muitas vezes, as técnicas de visualização de informação classificadas para dados quantitativos são usadas em uma abordagem com dados qualitativos. Em especial, se fez uso de algumas técnicas usadas na visualização de informação do tipo qualitativo e na dinâmica dessas informações ao longo do tempo.

Neste sentido, foram utilizadas diferentes tecnologias para criar ou estender diversas técnicas de visualização de informação, que permitiram avaliar a evolução de um conjunto de dados acadêmicos. A partir dessa evolução, diversos indícios sobre as informações podem ser levantados, facilitando a extração de conhecimento (descoberta de padrões ou de características) sobre esses dados.

O sistema proposto foi aplicado em uma base contendo os dados de grupos de pesquisa e professores de uma universidade no intervalo de 20 anos (1995-2015). Os resultados foram obtidos considerando os dados de grupos de pesquisa (em particular, as Produções Bibliográficas do grupo) e docentes (em particular, as Produções Bibliográficas e Trabalhos Orientados), fornecendo uma série de indícios a serem validados sobre estes dados posteriormente.

Cabe ressaltar, que a estratégia pode ser estendida para qualquer conjunto de características, individuais ou coletivas, que se deseja avaliar assim como foi feito nesse trabalho. Ao identificar quais técnicas de visualização de informação foram utilizadas nos trabalhos relacionados e em outras pesquisas estudadas, se selecionaram aquelas que podiam se adaptar aos dados do cenário de aplicação. Essas técnicas de visualização foram estendidas acrescentando dimensões temporais e dinamismo. Também foram selecionadas técnicas de visualização que maiormente são aplicadas aos dados quantitativos e depois de ser estudadas com

detalhe, se observou que podiam se adaptar também aos dados qualitativos deste trabalho, como os gráficos de barra ou circulares.

A partir desse conjunto de dados foi possível visualizar a evolução dos temas de pesquisa dos grupos no tempo. Também, foi possível observar individualmente o trabalho dos professores, se eles concentram sua atenção na publicação de trabalhos ou, pelo contrário, se seu foco de atenção é a orientação de trabalhos de alunos. Além de verificar os temas mais abordados em um determinado momento, a Nuvem de Palavras Animadas permite o acompanhamento da evolução de temas ao longo do tempo, a rede de publicações permite o acompanhamento dos principais temas de pesquisa do grupo e um gráfico circular permite acompanhar os principais temas desenvolvidos pelos professores individualmente.

No trabalho, foram demonstrados casos usando as técnicas de visualização desenvolvidas. Durante os testes foram realizados acompanhamentos dos grupos de pesquisa, porém o site também pode ser utilizado para acompanhar, individualmente, membros dos grupos. O acompanhamento de membros de um grupo de pesquisa pode revelar informações importantes sobre o engajamento do pesquisador nesse grupo.

De acordo com as categorias da ACM, a técnica de visualização criada permite observar em quais categorias estão localizadas as palavras-chave de um trabalho de pesquisa, se são recomendadas pela ACM e permite sugerir novas palavras-chave para um trabalho de pesquisa. Essa técnica ajuda ao professor a localizar melhor seu trabalho de pesquisa nas categorias corretas e ter melhores palavras-chave que descrevam corretamente o trabalho.

Este trabalho pode ser facilmente aplicado a outros grupos de pesquisadores sem problema, se o banco de dados cumpre com a estrutura definida neste trabalho, que é baseada na estrutura dos CV Lattes. Porém, uma limitação importante desta pesquisa é que os dados no banco não estão ao ano atual, sendo limitados até o início do ano de 2016. O processo de obter os dados dos CV Lattes dos professores e armazená-los no banco é muito complexo e tedioso, não foi feito um método automático para conseguir fazê-lo.

A última etapa da dissertação foi feita à distância. Uma vez na Venezuela, e tendo o *site online*, não foi necessário estar fisicamente no Brasil nem ter o site local para fazer os casos e experimentos e obter os resultados. Todas as pessoas que precisarem ter acesso à informação deste trabalho podem ter uma conta de acesso no *site*.

Como trabalho futuro, pretende-se criar uma ótima forma de obter os dados dos professores diretamente de seus CV Lattes ou da Plataforma Sucupira<sup>10</sup>, sem ter a necessidade de baixar os documentos XML e processá-los para armazenar os dados no banco de dados, podendo assim ter os dados atualizados à data. Uma possibilidade seria uma criação de um Serviço Web como uma fonte de dados.

Também, pretende-se desenvolver um módulo no *site* que permita ler os arquivos completos em formato PDF das publicações para interpretar o texto dos trabalhos e sugerir assim palavras-chave corretas aos professores.

Outra possibilidade seria o desenvolvimento de outras técnicas de visualização de informação do tipo texto onde poderiam ser observados diferentes dados das publicações bibliográficas dos membros dos grupos de pesquisa, como relação das publicações, comparação dos tipos de publicações, entre outros.

Uma visualização interessante que se pretende desenvolver é apresentar a concorrência das palavras-chave entre as publicações bibliográficas, assim como também a coautoria de autores.

Por fim, as técnicas de visualizações de informação propostas foram aplicadas em uma base de dados de um grupo relativamente pequeno de pesquisadores. Espera-se adaptar o trabalho a outros tipos de dados, tais como dados qualitativos e quantitativos de bases acadêmicas maiores, redes sociais, entre outras.

---

<sup>10</sup> <https://sucupira.capes.gov.br/sucupira/>

## 6.1 Publicações relacionadas

Até o presente momento, o trabalho de dissertação permitiu gerar duas publicações.

A primeira publicação apresentou genericamente algumas das ideias de visualização de dados contidas na proposta inicial do trabalho. Ela teve como contribuição uma revisão bibliográfica sobre a visualização de informação em mapas, as metodologias existentes para análise e os resultados preliminares obtidos através de uma Revisão Sistemática da Literatura. O artigo foi titulado “Visualização de dados em Sistemas de Informação Geográfica: uma revisão sistemática da literatura”, apresentado na Revista Espacios, aceito em abril de 2016 e publicado no Vol. 37 (Nº 19)<sup>11</sup>.

A segunda publicação estendeu o objetivo de pesquisa, não trabalhando só na visualização de informação em Sistemas de Informação Geográfica, senão trabalhando com diversas técnicas de visualização de informação, partindo para análise em títulos de artigos científicos, a fim de conseguir levantar novas hipóteses na pesquisa. O artigo foi denominado “*Animated Word Cloud to View and Extract Knowledge from Textual Information*” e foi aceito e apresentado durante as atividades do 22º Simpósio Brasileiro de Multimídia e Web (WebMedia 2016) realizado em novembro de 2016 em Teresina – PI<sup>12</sup>.

---

<sup>11</sup> <http://www.revistaespacios.com/a16v37n19/16371926.html>

<sup>12</sup> <http://dl.acm.org/citation.cfm?id=2976848>

---

## REFERÊNCIAS

- ARAÚJO, MATHEUS, et al. 2013. Métodos para análise de sentimentos no Twitter. In: Proceedings of the 19th Brazilian symposium on Multimedia and the Web (WebMedia'13).
- BARROS, D.; PRATES, R.; MELO-MINARDI, R. (2014). Uma investigação sobre as técnicas de visualização de informação quantitativas e qualitativas. Em: Proceedings of the 13th Brazilian Symposium on Human Factors in Computing Systems (pp. 397-400). Sociedade Brasileira de Computação.
- BASTIAN, M.; SEBASTIEN, H.; MATHIEU, J. "Gephi: an open source software for exploring and manipulating networks." ICWSM 8 (2009): 361-362.
- BHOWMICK, T. Building an exploratory visual analysis tool for qualitative researchers. In AutoCarto International Symposium on Automated Cartography (Vancouver, 2006), 1–13.
- CARD, K.; MACKINLAY, J.; SHNEIDERMAN, B. Readings in information visualization, using vision to think. 1999. San Francisco: Morgan Kaufmann. ISBN: 15 586 0533 9.
- CHEN, C. 2002. Visualization of knowledge structures. Handbook of software engineering and knowledge engineering, v. 2, p. 700.
- CHEN, C. 2005. Top 10 unsolved information visualization problems. Em: Computer Graphics and Applications, IEEE, vol. 25, no 4, p. 12-16.
- CHEN, C. Information visualization. Em: Information Visualization, 2002, vol. 1, no 1, p. 1.
- CRAFT, B.; CAIRNS, P. Directions for Methodological research in information visualization. Em: iV, 12th International Conference on Information Visualization, pp. 44-50, 2008, London: IEEE Computer Society Press.

- CRUZ, J. 2015. Utilizando títulos de artigos científicos na construção de redes semânticas para caracterizar áreas de pesquisa. Universidade Federal de Paraíba.
- CUI, W.; WU, Y.; LIU, S.; WEI, F.; ZHOU, M. X.; QU, H. 2010. Context preserving dynamic Nuvem de Palavras visualization. In: IEEE Pacific Visualization Symposium. pp. 121-128. URL: <http://research.microsoft.com/en-us/um/people/weiweicu/images/wordpv.pdf>
- DUARTE et al. 2014. WebMedia XX: Quem Somos e o Que Temos Feito nas Últimas Duas Décadas. In: WebMedia 2014. URL: <https://doi.org/10.1145/2664551.2664576>
- DÜRSTELER, J.C. 2004. Grafos. InfoVis, n.137. 5 Enero, 2004. URL: <http://www.infovis.net/printMag.php?num=137&lang=1>
- FAYYAD, U.; PIATETSKY-SHAPIO, G.; SMYTH, P. From data mining to knowledge discovery in databases. AI magazine, v. 17, n. 3, p. 37, 1996.
- FELDMAN, Ronen. 2013. Techniques and applications for sentiment analysis. Commun.of the ACM 56, 4, 82-89. DOI= <http://dx.doi.org/10.1145/2436256.2436274>
- FELDMAN, R; SANGER, J. 2006. The text mining handbook: Advanced approaches in analyzing unstructured data. Cambridge Univ. Press. ISBN: 13 978 0 521 83657 9.
- FIELDING, R. T. 2000. Architectural styles and the design of network-based software architectures. Tese de Doutorado. University of California, Irvine. URL: <http://www.ics.uci.edu/~fielding/pubs/dissertation/top.htm>
- FRY, B. (2007). Visualizing Data, Exploring and Explaining data with the processing environment. O'Reilly Media.
- GASPARINI, I.; KIMURA, M. H.; PIMENTA, M. S.; ALEGRE, P. Visualizando 15 anos de IHC. Em: IHC - Simpósio Brasileiro sobre Fatores Humanos em Sistemas Computacionais (Manaus, 2013), 238–247.
- GONZÁLEZ, Y. D.; ROMERO, Y. F. 2012. Patrón Modelo-Vista-Controlador. Revista Telem@tica, v. 11, n. 1, p. 47-57.

- HARDIN, M.; HOM, D.; PEREZ, R.; WILLIAMS, L. 2011. ¿Qué tabla o gráfico es el adecuado para usted?. Em: Tableau Software. Vol.6. URL: [https://www.tableau.com/sites/default/files/media/Whitepapers/which\\_chart\\_v6\\_es-final\\_0.pdf](https://www.tableau.com/sites/default/files/media/Whitepapers/which_chart_v6_es-final_0.pdf)
- HEARST, M. A. 2003. What is text mining? SIMS, UC Berkeley. URL: <http://people.ischool.berkeley.edu/~hearst/text-mining.html>
- HEER, J. 2010. A conversation with Jeff Heer, Martin Wattenberg, and Fernanda Viégas. Queue, v. 8, n. 3, 10 pp. DOI= <http://doi.acm.org/10.1145/1737923.1744741>
- HENDERSON, S.; SEGAL, E. H. Visualizing qualitative data in evaluation research. New Directions for Evaluation 2013, 139 (2013), 53–71.
- HOGUE, A.; KARGER, D. 2005. Automating the unwrapping of semantic content from the World Wide Web. In: Proc. 14th Int. Conf. on World Wide Web. ACM, p. 86-95. URL: <http://dl.acm.org/citation.cfm?id=1060762>
- KENDALL, J. E.; KENDALL K. E. Análisis e diseño de sistemas, 8va Edición. Pearson educación, 2011.
- MAIA et al. Analise da Rede de Colaboração do Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos: As Primeiras 30 Edições. 2012. In: Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos (SBRC). URL: [http://ce-resd.facom.ufms.br/sbrc/2012/STArtigosConvidados\\_2.pdf](http://ce-resd.facom.ufms.br/sbrc/2012/STArtigosConvidados_2.pdf)
- MARTIG, S. R.; CASTRO, S. M.; DI LUCA, S. Interacción en la visualización de información. In: IV Workshop de Investigadores en Ciencias de la Computación. 2002. URL: [http://sedici.unlp.edu.ar/bitstream/handle/10915/21829/Documento\\_completo.pdf?sequence=1](http://sedici.unlp.edu.ar/bitstream/handle/10915/21829/Documento_completo.pdf?sequence=1)
- MCNAUGHT, Carmel; LAM, Paul. 2010. Using Wordle as a supplementary research tool. In: The qualitative report. vol. 15, no 3, p. 630. URL: <http://www.nova.edu/ssss/QR/QR15-3/mcnaught.pdf>



- MONTERO, Y. H. 2006. Visualización e recuperación de información. II Encontro de Ciências e Tecnologias da Documentação e Informação.
- MURRAY, Daniel G. Tableau Your Data!: Fast and Easy Visual Analysis with Tableau Software. John Wiley & Sons, 2013.
- MURRAY, S. (2013). *Interactive data visualization for the Web*. Em: O'Reilly Media, Inc.
- NUALART V., J.; PÉREZ M., M.; WHITELAW, M. 2014. Cómo dibujamos textos: Revisión de propuestas de visualización e exploración textual. El profesional de la información, vol. 23, no 3, p. 221-235.
- OLMEDA, C. (2014). Visualización de información. Em: El profesional de la información, 2014, mayo-junio, v. 23, n. 3, pp. 213-219.
- OLMEDA, C. 2014. Visualización de información. El profesional de la información. Vol. 23, no 3, p. 213-9.
- PRESSMAN, R.; MAXIM, B. Software Engineering: A Practitioner's Approach. 8. ed. [S.I.]: McGraw-Hill Science/Engineering/Math, 2014. ISBN 0078022126.
- ROSSITE, D. 2016. Curso HTML, CSS and JavaScript. Em: The Hong Kong University of Science and Technology
- RUMBAUGH, J.; JACOBSON, G.; BOOCH, G. 2000. El lenguaje unificado de modelado: manual de referencia. Addison Wesley.
- SCHMULLER, J. 2004. Sams teach yourself UML in 24 hours. Sams publishing.
- SHNEIDERMAN, B. 1996. The eyes have it: A task by data type taxonomy for information visualizations. In: Proc. IEEE Symposium Visual Languages.pp. 336–343. DOI= <http://dx.doi.org/10.1109/VL.1996.545307>
- TORRES, D. La Visualización de la Información en el entorno de la Ciencia de la Información. Em: Universidad de Granada, 2010.

VIEGAS, F. B.; WATTENBERG, M.; FEINBERG, J. 2009. Participatory visualization with Wordle. *IEEE Transactions on Visualization and Computer Graphics*. vol. 15, no 6, p. 1137-1144. URL: <http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=5290722>

VIÉGAS, F.; WATTENBERG, M. 2008. Timelines tag clouds and the case for vernacular visualization. *Interactions*, v. 15, n. 4, p. 49-52, 2008.

VIEGAS, Fernanda; WATTENBERG, Martin; VAN HAM, Frank, KRISS; Jesse, MCKEON, Matt. 2007. Manyeyes: a *site* for visualization at internet scale. In: *IEEE Transactions on Visualization and Computer Graphics*, 13(6), 1121-1128.

WARE, C. (2012). *Information visualization: perception for design*. Em: Elsevier.

ZHU, B.; CHEN, H. *Information Visualization*. *Annual Review of Information Science and Technology*, 39, 139-177, 2005.

**APÊNDICE**

## APENDICE 1 - PLANO DE IMPLANTAÇÃO

Para a implantação do site, é preciso realizar o seguinte:

1. Realizar a instalação do Apache, MySQL e PHP
2. Realizar a instalação de GIT para baixar o repositório
3. Na pasta www do Apache se devem clonar os repositórios do projeto. A URL dos repositórios são: [https://felydl@bitbucket.org/cambios\\_cv/cv.git](https://felydl@bitbucket.org/cambios_cv/cv.git) e <https://betoharris@bitbucket.org/betoharris/cvfront.git>
4. Na raiz das pastas cv e cvfront se deve executar pela consola os seguintes comandos: `composer install`, `bower install`, `php app/console assets:install` e `php app/console assetic:dump --env=prod --no-debug`
5. Executar o site com o seguinte link: <http://localhost/cvfront/Web/login>

## APENDICE 2 - CONCEITOS E FERRAMENTAS

Nesse anexo são descritos os aspectos conceituais do sistema, permitindo conhecer e compreender as funcionalidades e comportamentos das tecnologias e ferramentas utilizadas no desenvolvimento do sistema proposto.

### Ferramentas

No processo de projeto e desenvolvimento do sistema, as seguintes ferramentas foram utilizadas:

TABELA 1 - FERRAMENTAS UTILIZADAS NO PROCESSO DE DESENVOLVIMENTO DO SISTEMA

Ferramenta	Descrição
NetBeans IDE 8.1 <sup>13</sup>	É um entorno de desenvolvimento integrado (IDE), composto por um grupo de ferramentas que são empacotadas em uma só.
MySQL Workbench 6.3 <sup>14</sup>	É uma ferramenta que permite modelar os diagramas da estrutura e criar bancos de dados MySQL.
SourceTree <sup>15</sup>	É uma ferramenta para gerenciar os sistemas de controle de versão em

---

<sup>13</sup> <https://netbeans.org/>

<sup>14</sup> <https://www.mysql.com/products/workbench/>

<sup>15</sup> <https://www.sourcetreeapp.com/>

	repositórios.
XAMPP <sup>16</sup>	XAMPP é uma distribuição de Apache totalmente gratuita e fácil de instalar que contém os pacotes para trabalhar com MariaDB, PHP e Perl. O pacote de instalação XAMPP foi projetado para ser extremamente fácil de instalar e usar.

Fonte: Elaborado pelo autor

## Tecnologias e linguagens

Nessa seção serão tratadas as tecnologias e linguagens que foram utilizadas no processo do projeto de desenvolvimento do sistema proposto.

### JQuery

jQuery<sup>17</sup> é uma biblioteca JavaScript que simplifica a forma de interagir com documentos HTML, manipular eventos, desenvolver animações e adicionar interação com a tecnologia AJAX a páginas Web. jQuery é software livre e *open source*, é duplamente licenciado sob a licença MIT e a Licença Pública Geral GNU v2, permitindo a sua utilização em projetos livres e privados. A biblioteca oferece funcionalidades que de outro modo exigiria muito mais código, ou seja, com as características desta biblioteca são obtidos grandes resultados em menos tempo e espaço.

---

<sup>16</sup> <https://www.apachefriends.org/es/index.html>

<sup>17</sup> <https://jquery.com/>

## Bootstrap

Bootstrap<sup>18</sup> é um *Framework* HTML, CSS e JavaScript ou conjunto de ferramentas de código aberto para a concepção de *sites* e aplicações Web. Ele é especialmente utilizado para o desenvolvimento de projetos responsivos e focado para dispositivos móveis na Web.

Bootstrap torna o desenvolvimento *front-end* Web mais rápido e fácil. Ele é feito para pessoas de todos os níveis e dispositivos de qualquer forma ou tamanho.

## HTML

*HyperText Markup Language*<sup>19</sup> (HTML) faz referência à linguagem de marcação para a elaboração de *sites* Web. É um padrão que define a estrutura e código do conteúdo de um *site* Web, como o texto, as imagens, os vídeos, os jogos e outros.

## JavaScript

JavaScript<sup>20</sup> é uma linguagem de programação interpretada. É definida como orientada a objetos, baseada em protótipos, imperativo e dinâmico. É usada principalmente do lado do cliente (*client-side*), implementado como parte de um navegador Web que permite melhorias na interface do usuário e *sites* Web dinâmicos, todos os navegadores modernos interpretam o código JavaScript integrado nos *sites* Web.

---

<sup>18</sup> <http://getbootstrap.com/>

<sup>19</sup> <https://www.w3.org/html/>

<sup>20</sup> <https://www.javascript.com/>

## CSS

Cascading Style Sheets<sup>21</sup> (CSS) é uma linguagem simples utilizada para adicionar estilos aos *sites* Web criados com HTML ou XML, por exemplo, tipos de letra, cores, espaço, etc.

## Apache

Apache<sup>22</sup> é um servidor Web HTTP responsável por aceitar pedidos de páginas (ou recursos em geral) que vêm de visitantes que acessam algum *site*, e gerenciar a entrega ou a recusa, de acordo com as políticas de segurança estabelecidas.

Apache é um projeto de código aberto e livre para uso, multiplataforma (com versões para todos os principais sistemas operacionais), muito robusto e conhecido pela sua segurança e desempenho. O objetivo deste projeto é fornecer um servidor seguro, eficiente e extensível que fornece serviços HTTP em sincronia com os padrões HTTP atuais.

## Symfony

Symfony<sup>23</sup> é um *framework* desenhado para otimizar o desenvolvimento dos sistemas Web, baseado no padrão Modelo Visão Controlador. O *framework* tem integrado ferramentas e classes que buscam reduzir o tempo de desenvolvimento de um sistema. Além disso, Symfony automatiza as tarefas comuns nos sistemas, permitindo ao desenvolvedor dedicar-se aos outros itens específicos. O objetivo de Symfony é apoiar aos desenvolvedores, para que não tenham que criar tudo do zero e possam gastar o seu tempo desenvolvendo coisas próprias do negócio.

---

<sup>21</sup> <https://www.w3.org/Style/CSS/>

<sup>22</sup> <https://httpd.apache.org/>

<sup>23</sup> <https://symfony.com/>



## Silex

Silex<sup>24</sup> é um *microframework* desenvolvido e compatível com a linguagem de desenvolvimento PHP. Fornece uma API (Application Programming Interface) fácil e intuitiva. É extensível, outro recurso muito útil, porque Silex utiliza o Pimple para a injeção de dependência, facilitando o uso de bibliotecas de terceiros e sendo fácil abstrair ainda mais o seu código. Silex também é fácil de provar porque utiliza recursos como o HttpKernel (raiz do Symfony) para *requests* e *responses*.

Este *microframework* é a ferramenta perfeita para iniciar a programação com Symfony, porque pode entender e praticar os conceitos principais dos seus componentes. Também pode ser utilizado para desenvolver com um paradigma orientado a objetos (OOP) de uso extensivo com Symfony, PHP e outros *frameworks* de última geração. Em suma, para projetos de complexidade média e baixa, Silex permite fazer um projeto bem feito, barato e em um período de tempo ideal. Além disso, o projeto terá um desempenho eficiente.

Silex foi o *microframework* utilizado para o desenho das visualizações estatísticas porque é compatível com PHP e Zingchart as principais bibliotecas utilizadas na geração e facilita o desenvolvimento.

## PHP

PHP<sup>25</sup> (Hypertext -processor) é uma linguagem de desenvolvimento do lado do servidor, originalmente concebida e usada para o desenvolvimento Web de conteúdo dinâmico. O código é interpretado por um servidor Web com um módulo do processador PHP que gera a página Web resultante. PHP evoluiu e agora também inclui uma interface de linha de comando que pode ser usada em aplicações gráficas independentes. Ele pode ser usado na maioria dos servidores Web, como em quase

---

<sup>24</sup> <http://silex.sensiolabs.org/>

<sup>25</sup> <https://secure.php.net/>

todos os sistemas operacionais e plataformas, sem nenhum custo. PHP é uma linguagem flexível, poderosa e de alto desempenho.

## MySQL

MySQL<sup>26</sup> é um sistema de gestão de bancos de dados relacionais, desenvolvido sob licença da Oracle Corporation, e é considerado o banco de dados de código aberto mais popular do mundo para ambientes de desenvolvimento Web<sup>27</sup>.

MySQL é amplamente utilizado em aplicações Web como Joomla, Wordpress, Drupal ou phpBB, em plataformas (Linux / Windows-Apache-MySQL-PHP / Perl / Python) e ferramentas de rastreamento de bugs como Bugzilla. Sua popularidade como uma aplicação Web está intimamente ligada ao PHP, que muitas vezes aparece em combinação com o MySQL.

MySQL é o banco de dados *open source* mais utilizado, que permite a aplicação de bases de dados confiáveis, de alto desempenho e facilmente escalável baseados na Web.

## Python

Python<sup>28</sup> é uma linguagem de programação multi-paradigma, porque suporta desenvolvimento em orientação a objetos, em programação imperativa e, em menor medida, na programação funcional. É uma linguagem interpretada, usa tipagem dinâmica e é multi-plataforma. É administrado pela Python Software Foundation. Tem uma licença de código aberto, chamado Python Software Foundation License, que é compatível com a *General Public License* (GNU).

---

<sup>26</sup> <https://www.oracle.com/es/products/mysql/overview/index.html>

<sup>27</sup> <http://db-engines.com/en/ranking>

<sup>28</sup> <https://www.python.org/>

### Padrão Modelo Visão Controlador (MVC)

O padrão MVC é um paradigma que divide as partes que compõem um aplicativo no Modelo, na Visão e nos Controladores, permitindo a implantação separadamente de cada elemento, garantindo a atualização e manutenção do sistema de forma fácil e em um pequeno espaço de tempo (GONZÁLEZ; ROMERO, 2012).

### Transferência de Estado Representacional (REST)

REST é uma interface entre os sistemas que usam HTTP para obter dados ou gerar operações com os dados em todos os formatos possíveis, tais como XML e JSON. É um estilo de arquitetura de software para sistemas hipermídia distribuídos, como a *World Wide Web*. Esta abordagem para o desenvolvimento de projetos e serviços Web foi definida por Roy Fielding<sup>29</sup>, o criador da especificação HTTP (FIELDING, 2000).

Entre as vantagens que REST fornece para o desenvolvimento estão:

1. Separação entre o Lado do Cliente e o Lado do Servidor: REST separa completamente a interface do usuário do servidor e o armazenamento dos dados. Melhora a portabilidade da interface para outras plataformas, aumenta a escalabilidade dos projetos e permite que os diferentes componentes dos desenvolvimentos possam evoluir de forma independente.
2. Visibilidade, confiabilidade e escalabilidade: A separação entre o cliente e o servidor permite que qualquer equipe de desenvolvimento possa escalar o produto sem problemas excessivos. Podem ser migrados para outros servidores ou fazer alterações no banco de dados, apenas se os pedidos dos dados forem enviados corretamente.

---

<sup>29</sup> <http://roy.gbiv.com/>

3. REST é sempre independente das plataformas ou linguagens: REST sempre se adapta ao tipo de sintaxe ou de plataformas que estão sendo trabalhadas, o que oferece a liberdade de mudar ou experimentar novos ambientes dentro do desenvolvimento. Com REST se pode ter servidores PHP, Java, Python ou Node.js. A única coisa que é essencial é que as respostas às solicitações sejam sempre feitas na língua do intercâmbio de informações usadas, geralmente XML ou JSON.

### Linguagem de Modelagem Unificada (UML)

UML (*Unified Modeling Language*) é um conjunto de ferramentas que permitem modelar (analisar e desenhar) sistemas. Também fornece algumas maneiras muito particulares aos programadores, desenvolvedores, analistas e designers de aplicações informáticas, as regras técnicas que permitam representar graficamente o comportamento e as estruturas que fazem parte de um sistema (RUMBAUGH; JACOBSON; BOOCH, 2000).

UML permite aos criadores de sistemas gerarem desenhos que capturam suas ideias de uma forma convencional e fácil de entender para comunicá-las aos outros (SCHMULLER, 2004).

UML através de seus elementos gráficos permite representar fluxos de trabalho de um grande sistema, facilitando:

- A comunicação entre os participantes (incluindo ferramentas) no desenvolvimento.
- Compreender as soluções (notação gráfica).
- A manutenção das soluções conceituais no tempo (Documentação).

## FosUserBundle

FOSUserBundle<sup>30</sup> é uma biblioteca de Symfony que permite a utilização de um subsistema de administração de usuários.

Este *bundle* fornece uma estrutura flexível de gerenciamento de usuários, permite registrar novos usuários, enviar uma mensagem de confirmação para os usuários recém-criados, editar o perfil dos usuários, adicionar campos no banco de dados para atribuir papéis, entre outras funções.

O componente FOSUserBundle fornece uma plataforma de segurança flexível que permite armazenar usuários desde a configuração, em um banco de dados, ou em outro lugar. FOSUserBundle é construído para ser rápido e fácil de armazenar os usuários em um banco de dados.

---

<sup>30</sup> <https://github.com/FriendsOfSymfony/FOSUserBundle>