
**Universidade Federal de Ouro Preto
Instituto de Ciências Exatas e Aplicadas
Curso Sistemas de Informação**

**Ferramenta de Suporte à Correção de
Trabalhos Acadêmicos em Formato Digital
através do Processamento de Linguagem
Natural e Comparação Textual**

Jáder Felipe Silva Duarte

**TRABALHO DE
CONCLUSÃO DE CURSO**

ORIENTAÇÃO:
Prof. Dr. Mateus Ferreira Satler

Agosto, 2017

João Monlevade/MG

Jáder Felipe Silva Duarte

**Ferramenta de Suporte à Correção de Trabalhos
Acadêmicos em Formato Digital através do Processamento de
Linguagem Natural e Comparação Textual**

Orientador: Prof. Dr. Mateus Ferreira Satler

Monografia apresentada ao
Curso de Sistemas de Informação do
Departamento de Ciências Exatas e
Aplicadas, como requisito parcial para
aprovação na Disciplina Trabalho de
Conclusão de Curso II.

Universidade Federal de Ouro Preto

João Monlevade

D812f Duarte, Jáder Felipe Silva.
Ferramenta de suporte à correção de trabalhos acadêmicos em formato digital através do processamento de linguagem natural e comparação textual [manuscrito] / Jáder Felipe Silva Duarte. - 2017.

86f.: il.: grafs; tabs.

Orientador: Prof. Dr. Mateus Ferreira Satler.

Monografia (Graduação). Universidade Federal de Ouro Preto. Instituto de Ciências Exatas e Aplicadas. Departamento de Computação e Sistemas de Informação.

1. Processamento de linguagem natural (Computação). 2. Ontologia (Computação). 3. Lógica difusa. 4. Portfólios em educação. 5. Inteligência artificial (Computação). I. Satler, Mateus Ferreira. II. Universidade Federal de Ouro Preto. III. Título.



UFOP
Universidade Federal
de Ouro Preto

UNIVERSIDADE FEDERAL DE OURO PRETO
INSTITUTO DE CIÊNCIAS EXATAS E APLICADAS
COLEGIADO DO CURSO DE SISTEMAS DE INFORMAÇÃO

ANEXO IV - Ata de Defesa

ATA DE DEFESA

Aos 25 dias do mês de agosto de 2017, às 10 horas e 30 minutos, na sala H 102 do Instituto de Ciências Exatas e Aplicadas, foi realizada a defesa de Monografia pelo aluno **Jáder Felipe Silva Duarte**, sendo a Comissão Examinadora constituída pelos professores: Prof. Dr. Mateus Ferreira Satler, Prof. MSc. Igor Muzetti Pereira e Prof. MSc. Elton Máximo Cardoso.

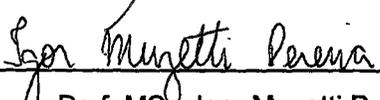
O candidato apresentou a monografia intitulada: "*Ferramenta de Suporte à Correção de Trabalhos Acadêmicos em Formato Digital através de Processamento de Linguagem Natural e Comparação Textual*". A comissão examinadora deliberou, por unanimidade, pela aprovação do candidato, com nota 9,5 (NOVE PONTOS E MEIO), concedendo-lhe o prazo de 15 dias para incorporação das alterações sugeridas ao texto final.

Na forma regulamentar, foi lavrada a presente ata que é assinada pelos membros da Comissão Examinadora e pelo graduando.

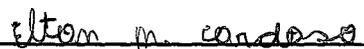
João Monlevade, 25 de agosto de 2017.



Prof. Dr. Mateus Ferreira Satler
Professor Orientador/Presidente



Prof. MSc. Igor Muzetti Pereira
Professor Convidado



Prof. MSc. Elton Máximo Cardoso
Professor Convidado



Jáder Felipe Silva Duarte
Graduando



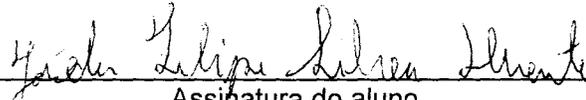
UFOP
Universidade Federal
de Ouro Preto

UNIVERSIDADE FEDERAL DE OURO PRETO
INSTITUTO DE CIÊNCIAS EXATAS E APLICADAS
COLEGIADO DO CURSO DE SISTEMAS DE INFORMAÇÃO

TERMO DE RESPONSABILIDADE

Eu, Jáder Felipe Silva Duarte, CPF: 090.061.206-18, declaro que o texto do trabalho de conclusão de curso intitulado *“Ferramenta de Suporte à Correção de Trabalhos Acadêmicos em Formato Digital através de Processamento de Linguagem Natural e Comparação Textual”* é de minha inteira responsabilidade e que não há utilização de texto, material fotográfico, código fonte de programa ou qualquer outro material pertencente a terceiros sem as devidas referências ou consentimento dos respectivos autores.

João Monlevade, 25 de agosto de 2017


Assinatura do aluno

DEDICATÓRIA

Dedico esta vitória aos meus pais, Sandra e HÉlbio que sempre me apoiaram e serviram de exemplo de persistência e fé, a minhas irmãs que sempre estiveram ao meu lado me incentivando, a Lorraine que sempre me motivou e apoiou, a meu amigo Igor que me incentivou na reta final e a todos os demais alunos, ex-alunos, amigos e colegas que de uma certa forma participaram da minha trajetória.

AGRADECIMENTOS

Agradeço a meus pais, Sandra E HÉlbio pelo apoio financeiro, moral e ético, além do amor incondicional. A minhas irmãs, Camila e Caroline pelo incentivo e pelas cobranças fundamentais. A minha namorada Lorraine pelo incentivo, amor e carinho. A Zezé, Dorinha e tia Cé pelo abrigo, acolhimento e os puxões de orelha. A meu orientador Mateus pelas correções pontuais e por acreditar em mim. A meu amigo Igor pela cobrança e incentivo na reta final. A meu cachorro Bob que sempre me recebe com festa mesmo que nem salgado eu tenha no bolso. Ao planeta Terra que continua girando.

Veni Codi Vici.

RESUMO

A massiva utilização dos meios digitais no ensino moderno proporciona comodidade e versatilidade ao ensino. Novos meios de aprendizagem são criados, tornando a educação mais democrática e acessível. Neste contexto, uma abordagem que vem crescendo no mundo acadêmico é a possibilidade de utilização do portfólio digital como ferramenta pedagógica. Porém, seu uso implica em certos desafios, como é o caso, por exemplo, do processo avaliativo do portfólio que demanda esforço significativo do avaliador. O objetivo deste trabalho é desenvolver uma ferramenta que auxilie neste processo avaliativo reduzindo a carga desta tarefa. Para tanto são utilizadas técnicas da “lógica fuzzy” aliadas ao poder representacional das “ontologias” de modo a representar semanticamente os portfólios através de um conjunto de relacionamentos entre pares de “conceitos”. Esses pares de conceitos são “palavras” que foram retiradas dos documentos pertencentes aos portfólios. Essas palavras passam por um “Processamento de Linguagem Natural” com o objetivo de prepara-las para o armazenamento numa estrutura que representará o “portfólio”. A saída da ferramenta é um relatório gerado a partir da comparação das ontologias fuzzy de dois portfólios: o portfólio de referência do tema analisado e o portfólio do aluno que será avaliado pelo professor.

Palavras-chave: Lógica Fuzzy, Ontologia, Ontologia Fuzzy, Processamento de Linguagem Natural.

ABSTRACT

The massive use of digital media in modern teaching provides convenience and versatility to teaching. New means of learning are created, making education more democratic and accessible. In this context, an approach that is growing in the academic world is the possibility of using the digital portfolio as a pedagogical tool. However, its use implies certain challenges, such as, for example, the evaluation process of the portfolio that requires significant effort by the evaluator. The objective of this work is to develop a tool that helps in this evaluation process by reducing the burden of this task. For that, fuzzy logic techniques are used, together with the representational power of "ontologies" in order to semantically represent portfolios through a set of relationships between pairs of "concepts". These pairs of concepts are "words" that have been taken from the documents belonging to the portfolios. These words undergo "Natural Language Processing" with the purpose of preparing them for storage in a structure that will represent the "portfolio". The output of the tool is a report generated from the comparison of the fuzzy ontologies of two portfolios: the reference portfolio of the analyzed theme and the portfolio of the student that will be evaluated by the teacher.

Keywords: Fuzzy Logic, Ontology, Fuzzy Ontology, Natural Language Processing.

LISTA DE FIGURAS

FIGURA 1: FUNÇÕES DE PERTINÊNCIA (GOMIDE; GUDWIN; TANSCHUIT, 1995).	28
FIGURA 2: TABELA DE PROJETOS DE PROCESSAMENTO DE LINGUAGENS NATURAIS (ALMEIDA; BAX, 2003).	33
FIGURA 3: EXEMPLO DE UMA ONTOLOGIA FUZZY. IMAGEM ADAPTADA DE SATLER (2011, P.47).	36
FIGURA 4: ESQUEMA DE APLICAÇÃO DAS TÉCNICAS DE PLN. IMAGEM ADAPTADA DE (SATLER, 2011, P.89)	38
FIGURA 5: REPRESENTAÇÃO DO MODELO DE SUPORTE PARA AVALIAÇÃO. IMAGEM GERADA PELO AUTOR.	50
FIGURA 6: DIAGRAMA DE CLASSE – CLASSE TERMO. IMAGEM GERADA PELO AUTOR.	71
FIGURA 7: DIAGRAMA DE CLASSE – CLASSE CONCEITO. IMAGEM GERADA PELO AUTOR.	72
FIGURA 8: DIAGRAMA DE CLASSE – CLASSE CONJUNTO FUZZY. IMAGEM GERADA PELO AUTOR.	72
FIGURA 9: DIAGRAMA DE CLASSE – CLASSE ONTOLOGIA FUZZY. IMAGEM GERADA PELO AUTOR.	73
FIGURA 10: DIAGRAMA DE CLASSE – CLASSE PORTFOLIO. IMAGEM GERADA PELO AUTOR.	74
FIGURA 11: DIAGRAMA DE CLASSE – CLASSE RELATÓRIO. IMAGEM GERADA PELO AUTOR.	75
FIGURA 12: JANELA PRINCIPAL DA FERRAMENTA DE AVALIAÇÃO DE PORTFÓLIO. IMAGEM GERADA PELO AUTOR.	76
FIGURA 13: DIAGRAMA DE CLASSE – CLASSE JANELA PRINCIPAL. IMAGEM GERADA PELO AUTOR.	76
FIGURA 14: DIAGRAMA DE CLASSE – CLASSE CONTROLADOR. IMAGEM GERADA PELO AUTOR.	78
FIGURA 15: RELATÓRIO DE AVALIAÇÃO DE PORTFÓLIO. IMAGEM GERADA PELO AUTOR.	81

FIGURA 16: GRÁFICO DE LINHAS COMPARATIVO DO TEMA LÓGICA FUZZY. IMAGEM GERADA PELO AUTOR.....	82
FIGURA 17: GRÁFICO DE LINHAS COMPARATIVO DO TEMA ROBÓTICA. IMAGEM GERADA PELO AUTOR.	84
FIGURA 18: NOTAS DOS ALUNOS DE BANCO DE DADOS NO TEMA BANCO DE DADOS MÓVEIS. IMAGEM GERADA PELO AUTOR.....	85

LISTA DE TABELAS

TABELA I: NOTAS DOS ALUNOS DE INTELIGÊNCIA ARTIFICIAL NO TEMA LÓGICA FUZZY.

TABELA GERADA PELO AUTOR.82

TABELA II: NOTAS DOS ALUNOS DE INTELIGÊNCIA ARTIFICIAL NO TEMA ROBÓTICA. TABELA

GERADA PELO AUTOR.83

TABELA III: NOTAS DOS ALUNOS DE BANCO DE DADOS NO TEMA BANCOS DE DADOS

MÓVEIS. TABELA GERADA PELO AUTOR.....84

LISTA DE FÓRMULAS

EQUAÇÃO 1: FUNÇÃO DE GRAU DE PERTINÊNCIA.....	27
EQUAÇÃO 2: FÓRMULA DA FREQUÊNCIA RELATIVA (WIVES, 1999, P.23).....	41
EQUAÇÃO 3: REPRESENTAÇÃO DA COLEÇÃO DE TUPLAS DE CONCEITO E PESO.	54
EQUAÇÃO 4: FÓRMULA DO CÁLCULO DO PESO DOS CONCEITOS NO PORTFÓLIO.	55
EQUAÇÃO 5: CÁLCULO DO GRAU DE SIMILARIDADE DOS CONCEITOS.	56
EQUAÇÃO 6: PADRONIZAÇÃO DO NÚMERO DE OCORRÊNCIAS DE UM CONCEITO EM UM DOCUMENTO.	56
EQUAÇÃO 7: APLICAÇÃO DA T-NORMA BD.....	57
EQUAÇÃO 8: CONJUNTO DE PORTFÓLIOS DOS ALUNOS.	58
EQUAÇÃO 9: FUNÇÃO DE PERTINÊNCIA DOS CONCEITOS.	59
EQUAÇÃO 10: FÓRMULA DE <i>JACCARD</i>	61

SUMÁRIO

1.	Introdução.....	16
1.1.	Justificativa.....	17
1.2.	Hipóteses.....	17
1.3.	Objetivo Geral e Objetivos Específicos.....	18
2.	Estado da Arte Referencial Teórico.....	19
2.1.	E-Portfólio.....	20
2.1.1.	Tipos básicos de portfólio.....	22
2.1.2.	Portfólio digital.....	24
2.2.	Lógica Fuzzy.....	25
2.2.1.	Conjuntos Fuzzy.....	27
2.2.2.	Operadores dos Conjuntos Fuzzy.....	29
2.2.3.	Inferência Fuzzy.....	29
2.3.	Ontologias.....	30
2.3.1.	Tipos de Ontologias.....	31
2.3.2.	Utilização de ontologias.....	32
2.3.3.	Ontologia Fuzzy.....	34
2.4.	Processamento de Linguagem Natural.....	37
2.4.1.	Níveis de PLN.....	40
2.4.2.	Frequência relativa de termos.....	41
2.5.	Linguagem de Programação JAVA.....	42
2.6.	Padrões de Projeto.....	44
2.6.1.	Padrão de Projeto MVC (<i>Model-View-Controller</i>).....	44
2.7.	Trabalhos Relacionados.....	45
2.7.1.	Um estudo sobre agrupamento de documentos textuais em processamento de informações não estruturadas usando técnicas de “Clustering”.....	46
2.7.2.	Enhancing Portfolio Assessment: An Application of Fuzzy Ontologies.....	47
3.	Proposta.....	48
3.1.	Definição do problema.....	48
3.2.	Modelo conceitual.....	49

3.2.1.	Processo de construção do portfólio	51
3.2.2.	Processo de geração do relatório.....	51
3.2.3.	Repositório dos Portfólios.....	52
3.3.	Estruturas do Modelo Conceitual	53
3.3.1.	Representação Conceitual do Portfólio	53
3.3.2.	Representação Conceitual do Relatório	57
3.3.3.	Coeficientes de similaridade.....	60
3.3.4.	Coeficiente de Jaccard.....	61
3.4.	Definição de capas	62
3.4.1.	Processo de criação do portfólio	62
3.4.2.	Processo de Geração do Relatório	66
4.	Implementação	69
4.1.	Camada Model	71
4.2.	Camada View	75
4.3.	Camada Controller	77
5.	Testes e Resultados	79
5.1.	Descrição do Experimento	79
5.2.	Resultados	80
6.	Conclusão e Trabalhos Futuros.....	86
	Referências	88

1. Introdução

Na era atual novos meios de comunicação surgiram e a forma com que nos comunicamos se alterou drasticamente. Neste novo cenário é vital a adaptação de todas as áreas para usufruir da melhor forma possível dos meios de comunicação modernos alcançando sucesso e excelência.

As tecnologias computacionais, que evoluíram muito nas últimas décadas, são responsáveis por grande parte do avanço dos novos meios de comunicação. Desta maneira, o acesso à informação é cada vez mais facilitado devido às novas formas de compartilhamento do conhecimento.

No meio acadêmico, houve uma grande melhora no acesso e no compartilhamento de artigos e monografias dos mais variados assuntos. A disseminação do conhecimento nas universidades e afins evoluiu e propiciou o aumento na produção de conhecimento por parte dos pesquisadores, graduandos, etc.

Não só as formas de acesso ao conhecimento, mas também as tecnologias educacionais sofreram mudanças significativas e abriram-se novos paradigmas. Técnicas como o *e-learning* vem contribuindo muito para a melhoria na educação surgindo como novos meios de aprendizagem para os alunos.

O *e-learning* é uma abordagem centrada na disponibilização das informações pelo computador através da internet, ou seja, possui um ambiente online, fato que torna o *e-learning* um importante propulsor da difusão do conhecimento pelo mundo, abrindo novos meios pelos quais a educação pode ser propagada. É uma abordagem muito utilizada ultimamente por sua versatilidade e disponibilidade em qualquer lugar e a qualquer hora, o que torna o ensino mais democrático e dinâmico.

As tecnologias atuais, além de abrirem novos paradigmas, abrem também novas perspectivas na educação. A aprendizagem colaborativa e a aprendizagem baseadas em problemas, são alguns dos exemplos de novas perspectivas de ensino que auxiliam na busca pela criação de ambientes de aprendizagem construtivistas, onde o aluno pode construir seu próprio conhecimento, compartilhando e enriquecendo-o com os demais alunos. Contudo, a aplicação destas novas ferramentas de ensino acabam acarretando em novos desafios para os professores

que pesquisam e tentam implantá-las, pois elas demandam novas habilidades e esforços adicionais destes educadores. (SATLER et al., 2011)

Neste contexto uma abordagem muito difundida e utilizada ultimamente é o e-portfólio, que nada mais é do que a utilização da metodologia de ensino através de portfólios digitais, com seus documentos e conteúdos diversos em formato digital. Em linhas gerais, a utilização de portfólios no processo de aprendizagem visa à melhora do pensamento reflexivo dos alunos, que podem avaliar seu próprio progresso analisando seus erros durante o processo. Desta forma, esta metodologia auxilia no processo de mudança da educação que busca novas formas para incentivar e melhorar a aprendizagem dos alunos, com o processo de aprendizagem centrado no aluno.

1.1. Justificativa

A abordagem portfólio vem ganhando cada vez mais adeptos e gerando novos desafios. O principal destes desafios está relacionado com o professor, que necessita de maior conhecimento e esforço para aplicar esta metodologia na sua disciplina ministrada. Além deste desafio, também podem ser citados: a dificuldade do aluno em absorver esta nova forma de obtenção de conhecimento e, por último, o esforço dedicado ao processo de avaliação destes portfólios.

Na literatura, as propostas encontradas para a avaliação dos portfólios são muito complexas ou por vezes imprecisas, principalmente quando as comparamos com a capacidade de julgamento do ser humano.

Nas pesquisas realizadas neste trabalho, observou-se que a utilização de tecnologias baseadas em análise semântica podem auxiliar no processo de avaliação de um portfólio. Desta maneira, o intuito deste trabalho é auxiliar o professor na avaliação dos portfólios gerados pelos alunos, possibilitando que ele possa usufruir mais deste importante mecanismo de aprendizagem.

1.2. Hipóteses

A utilização da ontologia fuzzy na representação de um portfólio digital de um aluno pode auxiliar no processo de avaliação deste portfólio. Essa avaliação pode ser realizada através da comparação desta ontologia fuzzy com outra ontologia fuzzy que representa um portfólio de referência.

1.3. Objetivo Geral e Objetivos Específicos

O objetivo geral deste trabalho é criar uma ferramenta para auxiliar professores na avaliação de portfólios digitais de alunos. Isto pode ser realizado através da criação de um relatório avaliativo para cada portfólio de aluno que serão representados por ontologias fuzzy.

Para atingir este objetivo, é necessário que um conjunto de objetivos específicos sejam alcançados. Dentre eles podemos enumerar:

- Buscar significados através de “conceitos”, que representam as palavras pertencentes aos documentos dos portfólios;
- Ponderar os “conceitos”, atribuindo-lhes um peso de representatividade no portfólio ao qual eles pertencem, que é calculado de acordo com a ocorrência destes conceitos dentro do referido portfólio;
- Gerar uma “ontologia fuzzy” desses “conceitos” através de um relacionamento entre cada conceito com os demais conceitos pertencentes aos portfólios, calculando e adicionando-os na “ontologia fuzzy” de cada portfólio. Esta estrutura representa todos os relacionamentos entre os “conceitos” pertencentes a um determinado portfólio.
- Por fim, gerar o “Relatório de Avaliação de Portfólios” através da comparação entre as ontologias fuzzy dos portfólios, que disporá de informações como: um conjunto de conceitos com maior valor de representatividade dentro do portfólio, os relacionamentos entre conceitos mais significativos de cada portfólio e o grau de compatibilidade entre os portfólios (calculado a partir das ontologias fuzzy).

E é este relatório que pode auxiliar o professor no processo avaliativo dos portfólios dos alunos. A próxima seção é dedicada ao Referencial Teórico deste trabalho. Na seção 3 serão abordados e explicitados o problema foco deste trabalho e o modelo utilizado como proposta de solução. Na seção 4 serão apresentados detalhes da criação da ferramenta para avaliação de portfólios. Na seção 5 serão apresentados os testes e resultados obtidos com o experimento utilizando a ferramenta criada. Na seção 6 são apresentadas a conclusão e os trabalhos futuros deste trabalho.

2. Estado da Arte Referencial Teórico

A rápida popularização da internet e dos meios de comunicação proporciona benefícios em escala global. Em segundos uma informação pode correr o mundo; as pessoas se comunicam instantaneamente independente da distância entre elas. Muitos benefícios poderiam ser citados aqui do advento da evolução dos meios de comunicação, mas com eles surgem também diversos desafios alavancados pelas diferentes necessidades da sociedade moderna, tachada hoje em dia como impaciente e acostumada com as facilidades que a internet provê. Estes desafios impulsionam a transformação dos sistemas de comunicação, que buscam entender a linguagem humana para atender as novas necessidades e expectativas da sociedade.

Trabalhos acadêmicos e artigos científicos podem ser acessados em qualquer lugar a qualquer hora graças à internet. Autores publicam suas pesquisas e trabalhos de diferentes áreas utilizando técnicas baseadas em ontologias, portfólios digitais, lógica fuzzy e linguagem natural, na busca por novos meios de resolução de problemas. A utilização dessas técnicas viabiliza a resolução de problemas que antes eram considerados intratáveis. Porém novos problemas e desafios surgem a cada dia.

Nos próximos tópicos serão abordadas as técnicas mencionadas, bem como algumas ideias de utilização conjunta destas técnicas na resolução do problema proposto neste trabalho.

2.1. E-Portfólio

O termo portfólio, de maneira simples, significa uma pasta de trabalhos, um conjunto de documentos relacionados a um projeto específico (ALVARENGA; ARAUJO, 2006a). São exemplos de portfólios: documentos e trabalhos realizados por um determinado profissional em uma empresa ou na sua carreira, documentos e informações contidas em uma pasta sobre dado projeto que está sendo executado em determinada empresa, exercícios e documentos gerados por um aluno em dado curso ou disciplina.

Na visão voltada para a área educacional, o portfólio tem ganhado importância nos últimos anos por ser um modelo pedagógico poderoso. Ele concentra-se na análise das contribuições realizadas pelos alunos refletindo a aprendizagem em algum assunto ou matéria. Essas contribuições são artefatos e documentos que o aluno entrega como parte de sua avaliação na disciplina em questão (SATLER et al., 2011). O processo de participação de um portfólio induz o aluno a uma auto avaliação oferecendo a possibilidade de maior fixação e ampliação de sua aprendizagem, além de torna-lo mais capaz de absorver corretamente o conteúdo e de gerar uma autocrítica saudável (ALVARENGA; ARAUJO, 2006a).

Harp e Huinsker (1977, p.224 apud ALVARENGA; ARAUJO, 2006) caracterizam portfólio como: “uma coletânea de trabalhos, que demonstram o crescimento, as crenças, as atitudes e o processo de aprendizagem de um aluno”. Desta maneira, segundo Alvarenga e Araújo (2006), em um portfólio devem ser contemplados planos e reflexões sobre os temas mais importantes tratados em sala, estudos de caso, relatórios e afins, entre outras ferramentas que devem constituir uma base para a avaliação evolutiva dos alunos em relação ao seu aprendizado. O principal ganho na utilização de portfólios é o desenvolvimento do pensamento reflexivo; por este motivo é muito interessante a sua utilização no meio acadêmico que busca novas formas de ensinar o aluno a pensar, a refletir.

Em contrapartida, temos diversas dificuldades e barreiras a serem transpostas para uma maior utilização dos portfólios na área acadêmica. Uma dessas barreiras é um maior esforço por parte do professor que deve elaborar o sistema de avaliação destes portfólios, bem como a especificação de critérios para a avaliação, a forma de aprendizagem dos objetos, o conteúdo, a interação do aluno e a avaliação das

atividades. Sendo de suma importância que o professor tenha um maior domínio dos meios digitais e ferramentas que possam auxiliar na avaliação e organização dos conteúdos dos portfólios, bem como outras habilidades que o professor deve desenvolver para usufruir desta metodologia da melhor forma possível. (SATLER et al., 2011)

Vale ressaltar que o portfólio é um instrumento de aprendizagem no caso acadêmico, tanto para o aluno como para o professor. Sendo assim avaliações por intermédio de portfólios demandam tempo, paciência e prática, além de um esforço extra do professor para avaliar quais objetos e trabalhos são mais importantes para o aprendizado e geram mais informações sobre a evolução dos alunos. (ALVARENGA; ARAUJO, 2006a)

A avaliação através de portfólios possui uma dificuldade singular por demandar demasiado esforço do professor na revisão dos artefatos dos alunos. A forma de avaliação destes artefatos pode ser executada observando-se aspectos diferentes, constituindo processos de avaliação diferentes, o que acarreta em um aumento na complexidade da avaliação.

Segundo Chen e Chen (2005) a avaliação de portfólios pode ser distinguida, basicamente, em dois processos de avaliação de aprendizagem: a avaliação somativa e a avaliação formativa. A avaliação somativa foca em analisar o resultado final do aluno verificando sua aprendizagem sem avaliar o progresso por ele obtido durante a execução das etapas; já na avaliação formativa o foco está na metodologia de aprendizado, onde são analisados todos os passos evolutivos do aluno até obter o resultado final, ou seja, todo o progresso do aluno durante o processo. É neste último modelo avaliativo que o portfólio se encaixa, por permitir uma visão global da evolução do aluno, o que auxilia na sua avaliação completa durante o processo de aprendizagem.

As avaliações formativas estão ganhando cada vez mais espaço por sua característica de avaliação do processo como um todo, podendo ser usada em diversas situações como, por exemplo: avaliação de desempenho com aplicações em empresas sobre empregados, de empresas sobre o próprio desempenho; na área educacional onde existe o maior número de pesquisas e aplicações dos portfólios, etc. Diversos trabalhos na literatura demonstram a aplicação de portfólios como metodologia de avaliação, um exemplo na educação pode ser verificado no trabalho

de Alvarenga e Araujo (2006b), bem como no trabalho de Gomes e outros (2010). Ambos defendem a utilização dos portfólios como metodologia de avaliação.

Os portfólios representam um papel importante na crescente utilização das avaliações formativas por permitirem a visualização da evolução do aluno, de todo o seu progresso durante uma disciplina, com a possibilidade do aluno poder visualizar seu progresso e refletir sobre seu próprio aprendizado.

A aplicação de portfólios no meio acadêmico é reforçada por Fourali (1997). O Autor cita a utilização da lógica fuzzy para a medição do nível educacional e ilustra como seus princípios poderiam auxiliar educadores na avaliação dos portfólios. Seus estudos foram voltados para a utilização desta técnica na avaliação de portfólios no Reino Unido, onde havia uma crise educacional. A mudança de metodologia de ensino proposta visava a melhoria na educação, uma espécie de evolução educacional baseada no êxito da utilização da lógica fuzzy em conjunto com os portfólios obtido em outros domínios, como por exemplo a indústria.

2.1.1. Tipos básicos de portfólio

O tipo de portfólio a ser utilizado está intimamente relacionado com o que se deseja coletar durante o processo, bem como a maneira que será realizada esta coleta. Mais de um método de portfólio pode ser utilizado ao mesmo tempo.

Danielson e Abruptin (1997 apud ALVARENGA; ARAUJO, 2006) caracterizaram pelo menos três tipos de portfólios básicos, o intitulado “display”, o “demonstrativo de trabalho” e o portfólio de “avaliação do trabalho por um determinado período”.

O chamado “display” tem sua utilização voltada para a documentação das atividades realizadas em sala de aula. Fotos de atividades executadas pelos alunos é um exemplo de uma forma valiosa de documentar visualmente a aprendizagem dos alunos, porém sem adicionar comentários que podem enriquecer e nortear o progresso, o desenvolvimento do aluno; assim a avaliação fica menos palpável, sendo recomendado que o professor ou o próprio aluno adicionem comentários, tornando a avaliação mais rica e diversificada e auxiliando no processo de aprendizagem como um todo.

Os portfólios chamados de “demonstrativos de trabalho” são aqueles que servem para mostrar os melhores artefatos gerados pelo aluno. Os documentos têm a possibilidade de serem escolhidos pelo professor, pelo aluno ou por ambos, usando como critério de escolha o melhor desempenho apresentado, ou seja, um trabalho que mais represente sua evolução. Um dos pontos negativos dessa abordagem é que a avaliação realizada pode não condizer com a realidade, uma vez que apenas se leva em consideração o melhor trabalho do estudante. Alguém que não participou do processo pode ter expectativas irreais sobre o desempenho dos estudantes. Além disso, informações sobre os caminhos percorridos pelo aluno para atingir aquele resultado são negligenciadas.

O terceiro tipo de portfólio, aquele que possui o objetivo de “avaliar o trabalho” de acordo com um tempo determinado, é dito ser o mais completo por demonstrar o ocorrido durante o processo de aprendizagem, permitindo que o professor e o aluno tenham uma visão clara do progresso e possibilitando determinar novos patamares a serem alcançados. Segundo Gronlund (1998 apud ALVARENGA; ARAUJO, 2006) o portfólio de avaliação do trabalho por um determinado período demonstra o que ocorreu durante o aprendizado de fatos, conceitos, procedimentos e atitudes nas atividades propostas ou no cotidiano. Alvarenga e Araújo (2006) reforçam que a avaliação cuidadosa dos trabalhos e sua relação com os objetivos pretendidos e com o próprio progresso fazem do portfólio de avaliação do trabalho uma ferramenta muito importante para a avaliação.

Além de escolher de forma adequada qual ou quais tipos de portfólios serão utilizados, é extremamente importante a elaboração do portfólio em si, de maneira a torná-lo mais eficaz, contemplando os aspectos necessários para uma avaliação adequada. Neste sentido, Gronlund (1998 apud ALVARENGA; ARAUJO, 2006) afirma que o pré-planejamento objetivando as bases curriculares de uma dada área, ou um esforço de áreas afins, tornam a coleta de documentação realmente válida e informativa para a avaliação. Dentro deste contexto, o autor cita quatro passos básicos na elaboração do portfólio:

- Determinar objetivos, competências e habilidades que serão desenvolvidas e avaliadas;
- Coleta de trabalhos com propósito e intenção;
- Determinar a organização dos trabalhos e o local de arquivamento;

- Seleção e avaliação dos trabalhos.

O primeiro dos passos é realizado antes do início das atividades a serem desenvolvidas pelos alunos, os demais podem ter a participação dos alunos.

A seguir será abordado o portfólio digital, muito utilizado hoje em dia na área educacional.

2.1.2. Portfólio digital

De maneira bem simples, um portfólio digital, ou e-portfólio, é uma versão eletrônica do portfólio. Muitos estudos deste formato de avaliação de aprendizado vêm sendo realizados por todo mundo, principalmente na Europa, onde se lançou o projeto COMENIUS entre 2005 e 2008, com o objetivo de desenvolvimento da carreira dos professores. Este projeto foi um esforço de diversas universidades e instituições de ensino de cinco países diferentes: Portugal, Itália, Hungria, Turquia e Polônia (LARANJEIRO; COSTA, 2008).

Laranjeiro e Costa (2008) afirmam que o foco principal do projeto foi a reflexão sobre as potencialidades e tecnologias digitais voltadas para o desenvolvimento profissional dos professores. Como resultado o projeto gerou um seminário internacional que tem por finalidade a disseminação do trabalho produzido e a possibilidade de usar este seminário como curso de formação de professores. Foram centenas de participantes de mais de treze países da união europeia que discutiram soluções e aprimoramentos avaliativos utilizando os portfólios digitais para fins educacionais.

O nome e-portfólio advém do fato de se utilizar recursos digitais no processo de avaliação por portfólios. São as ferramentas eletrônicas que estão relacionadas com os conteúdos abordados na disciplina ou matéria do curso em questão que auxiliam no processo de avaliação enriquecendo-os com vídeos, documentos, gráficos, imagens e programas de computador (SATLER et al., 2011).

A maioria das discussões educacionais atuais gira em torno da utilização dos portfólios digitais como metodologias para avaliação e organização dos estudos e como um método de aprendizado alternativo; porém o principal problema decorrente da utilização de tais métodos está na dificuldade imposta aos professores de estuda-

los e encontrar meios para utilizá-los de forma a obter maior proveito possível e que realmente os alunos aprendam a utilizá-los podendo refletir sobre seus próprios resultados e evoluir de forma saudável.

Na próxima seção será abordada a lógica fuzzy e suas contribuições para este trabalho.

2.2. Lógica Fuzzy

O termo Fuzzy, em inglês, tem muitos significados e em português a tradução é dada nos conceitos “difuso”, “nebuloso”, “incerto”. Sendo assim a lógica fuzzy trata da incerteza nos sistemas computacionais. A representação de certas informações pode gerar ambiguidade, como o exemplo clássico do copo: “O copo está meio cheio, ou meio vazio?”. São necessários padrões para a representação de tais valores. Para tratar tais imprecisões nos dados é que a lógica fuzzy existe.

A matemática é uma ciência de natureza exata. Na lógica comum existem valores precisos ou exatos que representam dados e informações. Estes valores são os valores binários que são representados por 0's e 1's, ou falso e verdadeiro, ou ainda podem ser interpretados como ligado e desligado. Suas proposições assumem apenas um destes valores: verdadeiro ou falso. Já a lógica fuzzy é um ramo matemático que suporta os modos de raciocínio aproximados, ou numa visão mais simplista, a melhor representação do raciocínio humano, que é essencialmente parcial e aproximado (SOUZA, 2009, p.49-50).

Uma proposição fuzzy assume não apenas um valor e sim um conjunto de valores aproximados, sendo eles os chamados conjuntos fuzzy, descritos linguisticamente e.g.: “muito quente”, “quente”, “nem quente nem frio”, “frio”, “muito frio”, e os termos linguísticos, que são um subconjunto fuzzy. Destes termos linguísticos obtemos uma definição que é a variável linguística, cuja finalidade é capturar os valores dentro de um conjunto de termos linguísticos, as palavras ou frases. Esta variável linguística assume instancias linguísticas e não numéricas como as variáveis numéricas. Por exemplo: uma variável linguística *Altura* pode assumir como valor um dos membros do seguinte conjunto: {baixa, média, alta}. A atribuição

de significado para os termos linguísticos se dá pela associação deles a um conjunto fuzzy definido em um universo de discurso comum (GOMIDE; GUDWIN, 1994).

Segundo Zadeh (1965) a motivação para a criação da lógica fuzzy vem do fato de que as classes de objetos do mundo real não possuem critérios específicos de pertinência, como por exemplo, a classe de pessoas altas. Cada pessoa possui uma visão diferente quanto a isso. Uma pessoa com 1,70m pode ser considerada alta para um grupo de pessoas, mas de estatura mediana ou baixa para outras; o mesmo ocorre com a classe de pessoas bonitas, a classe de pessoas jovens, a classe de temperaturas agradáveis, etc. Todas essas medidas são baseadas em visões subjetivas, o que significa que diferentes indivíduos podem ter opiniões discrepantes acerca de um mesmo assunto ou questão. A matemática tradicional não trata de tais questões, o que demandou a criação de um mecanismo para tal.

A lógica fuzzy é a base para a resolução de problemas considerados impossíveis de serem tratados com as técnicas clássicas. Ela possibilita o desenvolvimento de algoritmos e métodos de modelagem e controle dos processos, o que permite a redução substancial da complexidade dos projetos e suas implementações, fazendo da lógica fuzzy uma importante ferramenta para uma grande gama de aplicações em diversas áreas (GOMIDE; GUDWIN, 1994).

A lógica fuzzy se difere da lógica convencional em diversos aspectos. Gomide, Gudwin e Tanscheit (1995) sumarizaram as diferenças básicas, como pode ser visto a seguir:

- Na lógica fuzzy os predicados são nebulosos e nos sistemas lógicos binários os predicados são exatos;
- Na lógica clássica o modificador de predicados mais utilizado é a negação, enquanto que na lógica fuzzy admite-se diversos, tais como: muito, mais, menos, etc.;
- Os quantificadores na lógica comum são apenas o existencial e o universal; já na lógica fuzzy existe uma grande variedade deles, tais como: pouco, vários, usualmente, etc.

As diferenças básicas que a lógica fuzzy demonstra em relação à lógica convencional lhe conferem um poder de representatividade e mesmo de simplicidade tão grandes que sua utilização se torna obrigatória em diversos problemas em áreas

muito distintas, por reduzir a complexidade de alguns projetos com problemas intratáveis a tal ponto que os faz possíveis de serem realizados.

Na próxima subseção serão abordados os conjuntos fuzzy e será feita uma diferenciação dos conjuntos matemáticos da teoria clássica de conjuntos.

2.2.1. Conjuntos Fuzzy

Para falarmos de conjuntos fuzzy, antes precisamos falar de conjuntos matemáticos. Portanto, nos próximos parágrafos serão abordados os conceitos gerais referentes a teoria dos conjuntos.

Segundo Souza (2009, p.52), um elemento pertence ou não a uma determinada classe, de acordo com a teoria clássica dos conjuntos. Assim dado um universo de discurso arbitrário U e um elemento particular $z \in U$, temos que o grau de pertinência $\mu_A(z)$ com relação a um conjunto $A \subseteq U$ é tido como:

$$\mu_A(z) = \begin{cases} 1 & \text{se } z \in A \\ 0 & \text{se } z \notin A \end{cases}$$

Equação 1: Função de Grau de Pertinência.

Na função 1 é evidenciado que o grau de pertinência de um elemento z em relação a um conjunto A só pode assumir um dos valores: 1 se o elemento pertencer à A e 0, caso ele não pertença. Desta forma podemos verificar que a função característica do grau de pertinência, de acordo com a teoria clássica de conjuntos é: $\mu_A(z) : U \rightarrow \{1,0\}$. Reforçando que o grau de pertinência só assume valores discretos.

A manipulação de dados com erros limitados é um exemplo de uso desta ideia de forma generalizada. Nela os números que estiverem dentro do percentual de erro terão o fator de pertinência 1 e os demais terão o fator de pertinência 0. Este fato é demonstrado na figura 1(a) a seguir. No caso preciso, o fator de pertinência igual a 1 é dado somente para os números de igual valor, ou seja todos valorados em 1 e para os demais o fator correspondente é 0, como evidenciado na figura 1(b) (GOMIDE; GUDWIN; TANSCHKEIT, 1995).

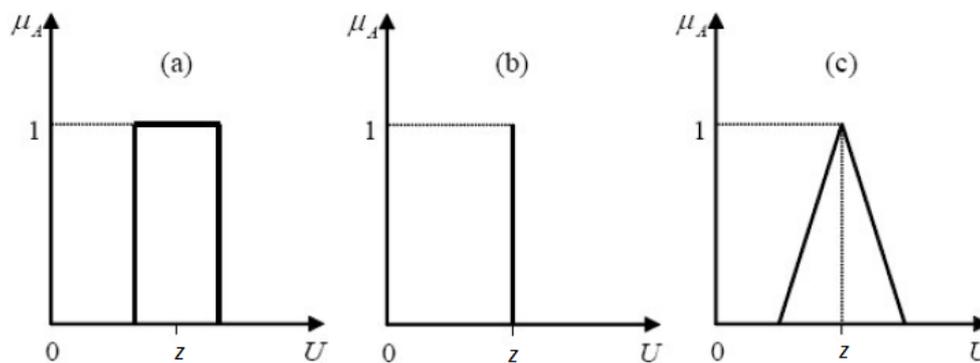


Figura 1: Funções de pertinência (GOMIDE; GUDWIN; TANSCHKEIT, 1995).

Uma caracterização mais abrangente foi proposta em 1965, por um autor que é considerado o pai da lógica fuzzy. Lotfali Askar Zadeh (1965) então professor na Universidade da Califórnia em Berkeley. Ele deu uma visão mais ampla de conjuntos ao sugerir que alguns elementos são mais pertencentes ao conjunto do que outros. Assim o fator de pertinência que na lógica clássica dos conjuntos podia assumir apenas 1 ou 0, agora teria a possibilidade de assumir quaisquer valores compreendidos no intervalo de $[0,1]$, onde 1 significa total pertinência e 0 representaria a pertinência nula. Esta ideia possibilitou o aumento de expressividade da função característica como evidenciado na figura 1(c). Nela observamos uma curva de pertinência triangular, onde números que estiverem próximos ao valor de z possuem uma maior identificação com o conceito representado por z .

Formalmente temos que um conjunto nebuloso A do universo de discurso U que pode ser contínuo ou discreto, é definido pela função de pertinência $\mu_A : U \rightarrow [0,1]$, onde esta função associa a cada elemento z de U o valor $\mu_A(z)$ que indica o grau de compatibilidade entre z e o conceito expressado por A onde (SANDRI; CORREA, 2006):

- $\mu_A(z) = 1$ indica que z é compatível com A ;
- $\mu_A(z) = 0$ indica que z é incompatível com A ;
- $0 < \mu_A(z) < 1$ indica que z é parcialmente compatível com A , com um grau de $\mu_A(z)$.

Dadas as diferenças conceituais com os conjuntos fuzzy, são necessárias adaptações na lógica comum para analisar os dados gerados por eles, bem como

propriedades e padronizações que transformem os dados gerados em informações úteis. Assim criou-se a lógica fuzzy para atender as necessidades representativas dos conjuntos fuzzy, uma lógica nova com regras um pouco diferentes nas operações a serem realizadas nos conjuntos, propriedades matemáticas, entre outras.

2.2.2. Operadores dos Conjuntos Fuzzy

De maneira análoga aos conjuntos ordinários, as propriedades algébricas de involução, idempotência, comutatividade, associatividade, distributividade e a lei transitiva são aplicadas nos conjuntos fuzzy segundo GOMIDE e GUDWIN (1994). O mesmo acontece para as operações de intersecção e união com o chamado conjunto vazio \emptyset e com o universo U . Agora as operações $A \cap A' \neq \emptyset$ e $A \cup A' \neq U$ ocorrem nos conjuntos fuzzy, enquanto que nos conjuntos ordinários temos: $A \cap A' = \emptyset$ e $A \cup A' = U$.

As operações de União, Intersecção e Complemento possuem particularidades no uso em conjuntos nebulosos, como fica evidenciado nos trabalhos de Gomide e Gudwin (1994), Sandri e Correa (2006) e Souza (2009). Para a intersecção foi criado um conjunto de normas, as chamadas *t-normas*, para a união de conjuntos criou-se as *t-Conormas* ou *s-normas*. Algumas destas normas e regras foram criadas por Zadeh e complementadas por outros autores ao longo do tempo. No caso da propriedade “complemento”, pelo menos duas regras devem ser satisfeitas, de acordo com Moraes (2008): a monotonicidade e a condição de contorno. Neste caso, define-se o complemento de A como: $\mu_{A'}(z) = 1 - \mu_A(z)$.

2.2.3. Inferência Fuzzy

Um sistema de inferência fuzzy, nada mais é do que um sistema que gera conclusões a partir de uma base de regras fuzzy, do tipo “se - então”, ou seja, de condição e consequência. Este sistema de inferência ocorre após a transformação dos dados para a representação fuzzy, sendo assim descrito: os valores de entrada sofrem uma normalização e posterior conversão para a representação fuzzy, método

conhecido como fuzzyficação das entradas. Paralelo a isso, ocorre à execução das regras do sistema fuzzy e para cada variável do sistema é produzida uma região fuzzy. Este passo é conhecido como agregação. Depois disso ocorre a defuzzificação, ou seja, as regiões fuzzy são convertidas em valores reais, ou valores “crisp”, obtendo-se o valor de saída de cada variável do sistema (Souza 2009, p.51).

Como a lógica fuzzy trabalha sobre termos linguísticos, que são conhecidamente mais próximos do pensamento humano; é necessária a conversão dos valores de entrada para valores fuzzy e, posteriormente, em valores reais, que possam ser manipulados computacionalmente.

A lógica fuzzy, em conjunto com estratégias de representação do conhecimento como as ontologias, pode gerar boas ferramentas para a manipulação de termos linguísticos e muitas outras finalidades. Desta forma, na próxima seção, as ontologias serão abordadas.

2.3. Ontologias

A W3C, World Wide Web Consortium, define ontologia como “termos utilizados na descrição e na recuperação de uma área do conhecimento” (INE, 2016a). Num contexto mais filosófico, uma ontologia é uma ciência que estuda os tipos e estruturas de objetos, propriedades, eventos, processos e relações que se referem a um objeto específico de cada área (BEPPLER, 2008, p.44). Na área de sistemas de informação e inteligência artificial, ontologia é vista como uma descrição de parte do conhecimento de um programa. Desta forma uma ontologia deve ser expressada formalmente para facilitar o processamento, de forma a se tornar adequada para programas de computador (BEPPLER, 2008, p.44-45).

De uma forma mais simplificada, ontologias são estruturas que organizam informações através de conceitos e seus relacionamentos. Uma definição é dada por Gruber (1993a apud BEPPLER, 2008, p.44), que diz o seguinte: “Uma ontologia é uma especificação formal e explícita de uma conceitualização compartilhada.”, podendo ser melhor interpretada como uma visão simplificada do mundo em questão, sendo representado por conjuntos de conceitos e suas relações, definições, propriedades e restrições.

As ontologias definem regras de combinação entre termos e suas relações dentro de um domínio do conhecimento. São descritas em forma de axiomas, ou seja, na forma de leis ou princípios que auxiliam na especificação do relacionamento entre conceitos (DUARTE; FALBO, 2000). A utilização dos axiomas é feita na organização do conhecimento destes domínios, o que possibilita a representação adequada destes conhecimentos propiciando o desenvolvimento de sistemas mais inteligentes e com melhores funcionalidades. Através destes axiomas são definidos conceitos inter-relacionados que representam o conhecimento de termos específicos de um dado domínio sobre um documento.

Atualmente as ontologias têm recebido uma maior atenção no que tange a organização dos dados, devido à facilidade em se trabalhar com esta metodologia para este fim. Outras técnicas como os Tesouros e as redes semânticas, podem substituir ou auxiliar na criação de uma ontologia, pois também organizam os dados a partir de conceitos e seus relacionamentos. A aplicação da ontologia em algumas áreas ainda não foi completamente estudada; sendo assim, não é raro encontrarmos contradições e ambiguidades que dificultam na escolha e na correta utilização de ontologias. As técnicas de manipulação destas ontologias no processo de organização da informação, além da grande gama de opções entre os tipos de ontologia deixam o processo mais nebuloso e menos natural (ALMEIDA; BAX, 2003). No que diz respeito aos tipos de ontologias, existem diferentes interpretações e visões por parte dos autores. Na próxima subseção esse tema será tratado com mais detalhes.

2.3.1. Tipos de Ontologias

As ontologias dificilmente apresentam a mesma estrutura. Em contrapartida apresentam componentes comuns e características similares. Beppler (2008, p.45-46) cita uma divisão das ontologias com direcionamento para criação de sistemas baseados em conhecimento, o que vem de encontro aos propósitos deste trabalho. A divisão é dada em três partes:

- **Ontologias de Domínio:** possuem a representação de conhecimento de um domínio específico. Exemplos: engenharia mecânica, comércio eletrônico, biologia, eletrônica, etc.;

- **Ontologias de senso comum ou genéricas:** também conhecida como ontologia de alto nível, este tipo de ontologia não depende de domínio específico e guarda informações gerais sobre o mundo, noções básicas e conceitos, tais como: tempo, espaço, estado, ação, etc.;
- **Ontologias representacionais:** são mais voltadas a visões filosóficas e lógicas ao invés de serem voltadas a aplicações como as demais. Têm a possibilidade de aplicação em âmbito geral por possuírem uma descrição de estruturas conceituais e meta-estruturas.

Na divisão realizada por Beppler, é clara a distinção das ontologias de acordo com o desenvolvimento de sistemas baseados em conhecimento. Outros autores também geraram diferentes divisões de tipos de ontologias, cada um com uma abordagem ou foco diferentes; em alguns casos podemos encontrar muitas semelhanças entre os tipos propostos pelos autores. Almeida e Bax (2003), por exemplo, geraram uma tabela que apresenta a classificação dos tipos de ontologias de acordo com as seguintes abordagens: “Quanto à função”, “Quanto ao grau de formalismo”, “Quanto à aplicação”, “Quanto à estrutura” e “Quanto ao conteúdo”. Essa classificação é mais específica do que a proposta por Beppler, pois para cada ramo há uma abordagem que pode ser seguida para um melhor entendimento e facilidade do trabalho com a ontologia, exatamente porque a divisão é mais específica. Contudo é necessário entender muito bem essa divisão proposta pelos autores a fim de evitar erros durante a geração da ontologia.

Uma outra possibilidade à criação de ontologias é a utilização de outras ontologias para este fim, o que auxilia no processo de configuração e diminui os erros na construção. Ou ainda pode-se utilizar ontologias presentes em repositórios disponíveis na internet. Essas ontologias podem ser adaptadas e utilizadas de acordo com a necessidade.

2.3.2. Utilização de ontologias

As ontologias podem ser criadas para atender diferentes necessidades. Noy e McGuinness (apud BEPPLER, 2008, p.48) citam cinco razões pelas quais se

explicaria a criação de ontologias: (a) compartilhamento de uma estrutura de informação entre os programas e pessoas e a possibilidade de compreensão desta estrutura; (b) reuso do conhecimento de um domínio; (c) transparência das definições de um domínio; (d) dissociação dos conhecimentos operacional e de domínio; (e) análise do conhecimento do domínio. A finalidade do uso de ontologias é vasta. Permite diversos benefícios que vão desde a criação de uma estrutura para a reutilização dos conhecimentos inerentes a um dado domínio, até uma melhor organização e compartilhamento destes conhecimentos para consultas e análises.

Em diversas áreas se observou a importância da utilização das ontologias. Alguns exemplos são: engenharia e representação do conhecimento, modelagem de banco de dados, integração, recuperação e extração de informações, desenvolvimento de sistemas, entre outras. Almeida e Bax (2003) realizaram um estudo sobre a utilização das ontologias em diversos domínios, tais como: “a gestão do conhecimento”, “comércio eletrônico”, “processamento de linguagens naturais”, “recuperação da informação na web”, entre outros. Na figura 2 são apresentados alguns exemplos de projetos de ontologias desenvolvidas para o processamento de linguagens naturais, com uma breve descrição.

Projetos relacionados ao processamento de linguagens naturais

Projeto	Breve descrição
<i>Oncoterm</i>	Facilita a tradução de textos médicos sobre oncologia mediante uma ontologia baseada em textos especializados e dicionários médicos; os conceitos são organizados em categorias e representados por esquemas (Moreno & Hernández, 2000).
<i>Gazelle</i>	Traduz textos japoneses, árabes e espanhóis para o inglês; inclui processamento e análise semântica das línguas, geração de sentenças em inglês, construção de ontologias interlíngua e criação de léxicos para japonês, árabe, espanhol e inglês (Germann, 1998).
<i>Mikrokosmos</i>	Mecanismo em que textos em linguagem natural são traduzidos para textos no formato interlíngua (linguagem neutra, chamada TMR- <i>text meaning representation</i>); a ligação da ontologia com o TMR é feita por meio de um léxico, no qual os significados dos itens são definidos por conceitos ontológicos (Beale, Nirenburg & Mahesh, 1995).
<i>PANGLOSS</i>	Possui um sistema baseado em exemplos e um sistema de transferência léxica que trabalham em paralelo propondo traduções das entradas; o resultado ou tradução final é selecionada por um modelo estatístico (Frederking, 1994)
<i>KPML (Komet-Penman MultiLingual)</i>	Desenvolve gramáticas e gera linguagens naturais; oferece um ambiente de desenvolvimento gráfico para a construção, manutenção e uso de gramáticas para diversas línguas (inglês, alemão, holandês, chinês, espanhol, russo, búlgaro e checo) (Bateman, 1996).
<i>Ontogeneration</i>	Gera textos em espanhol no domínio da química e utiliza linguagens naturais para responder a consultas sobre grupos, elementos e propriedades químicas; utiliza uma ontologia da química (<i>Chemicals</i>), uma ontologia lingüística e uma gramática em espanhol (Aguado <i>et alii</i> , 1998).
<i>Penman</i>	Gera sentenças em linguagem natural a partir de uma entrada não lingüística, aceita várias notações na entrada e foi projetado para o uso por pessoas com vários graus de sofisticação lingüística e computacional (Teich & Bateman, 1993).
<i>TechDoc</i>	Gera documentos técnicos multilingüísticos (inglês, alemão e francês) a partir de uma representação independente construída pela análise comparativa de partes dos manuais técnicos e lingüísticos disponíveis em diferentes línguas (Rösner & Stede, 1994).

Figura 2: Tabela de projetos de Processamento de Linguagens Naturais (ALMEIDA; BAX, 2003).

Os projetos evidenciados na figura 2 são exemplos de aplicações de ontologias no processamento de linguagem natural. Alguns destes projetos possuem como proposta a tradução ou geração de textos de diferentes idiomas, dicionários inteligentes para termos da medicina, dentre outras finalidades. Estes trabalhos, utilizam a linguagem natural para diferentes fins, gerando ontologias interessantes em domínios diferentes.

As ontologias, de um modo geral, representam o conhecimento inerente ao domínio ao qual foram submetidas. Na prática, a análise de alguns documentos científicos e posteriormente sua categorização em um domínio ou área de conhecimento é complexo. As palavras extraídas destes documentos não podem ser tratadas da mesma forma. Algumas palavras são mais significativas que outras na composição do significado do documento. Desta maneira, é difícil definir se um documento pertence integralmente à determinada área. Domínios possuem definições de conceitualização diferentes, especialistas de cada área possuem visões diferentes sobre estes domínios e suas alçadas. A ambiguidade, desta forma, se torna inevitável tanto na definição dos conceitos do domínio quanto na definição da relação entre estes conceitos, e mais ainda quanto ao peso que estas palavras têm na composição do significado do documento.

Tais problemas são tratados em diversos estudos. Como exemplo, podemos citar o trabalho de Satler (2011) que propôs a utilização da lógica fuzzy aliada às ontologias como forma de solucionar a manipulação do conhecimento impreciso e ambíguo. Tal técnica, chamada de ontologia fuzzy, será tratada neste artigo e melhor discutida na próxima seção.

2.3.3. Ontologia Fuzzy

Diversas técnicas vêm sendo testadas e novas formas de resolução de problemas vão sendo desenvolvidas, principalmente com a utilização conjunta de técnicas bem conhecidas, como é o caso das chamadas Ontologias Fuzzy. Em termos gerais, pode-se entender uma Ontologia Fuzzy como sendo a aplicação da lógica fuzzy na modelagem de uma ontologia.

A incorporação da lógica fuzzy na construção de uma ontologia lhe confere a capacidade de manipulação do conhecimento impreciso e eficiência comprovada na recuperação e representação de textos e multimídia (Zhai et al. 2007 apud SATLER, 2011, p.46). A importância de inserirmos a lógica fuzzy nos processos de uma ontologia está no fato dela proporcionar aumento do poder de representação da ontologia, o que a torna mais robusta e mais próxima da representação do pensamento humano.

A ontologia fuzzy pode ser encontrada em diversos trabalhos na literatura atual com abordagens diferentes. Um exemplo é o trabalho de Pereira, Ricarte e Gomide (2005) onde é apresentado um modelo de ontologia relacional fuzzy para sistemas de recuperação de informação com ênfase em algoritmos de busca. Outro exemplo é o trabalho de Satler (2011), que gerou perfis de usuário baseados em ontologias fuzzy para melhorar a filtragem de informação.

A junção da ontologia com a lógica fuzzy pode ser feita de formas diferentes, com níveis diferentes. Neste sentido, Yaguinuma (2007 apud AYRES, 2012, p.54) cita algumas das principais abordagens de ontologia fuzzy que podem ser encontradas na literatura, a saber: “taxonomias fuzzy”, “ontologias contendo classes e relacionamentos fuzzy” e “ontologias contendo hierarquia fuzzy de classes e relacionamentos”.

Destas abordagens citadas, a mais simples é a de “taxonomia fuzzy”, onde os termos são relacionados por generalização ou especificação, possuindo graus de pertinência entre eles. Já as “ontologias com classes e relacionamentos fuzzy” são mais completas que as anteriores, que apesar de possuírem algumas similaridades nas instâncias, permitem o inter-relacionamento destas instâncias, fato que não ocorre na taxonomia fuzzy. A abordagem das “ontologias contendo hierarquias fuzzy de classes e relacionamentos” é mais abrangente e completa que as duas anteriores por permitir também a existência de “graus de pertinência” entre as classes e subclasses, tornando-a mais robusta e com maiores possibilidades de utilização (AYRES, 2012).

Segundo Satler (2011, p.46-47), a ontologia fuzzy se baseia na ideia de definição de graus de pertinência da lógica fuzzy, nos relacionamentos entre os termos ou conceitos que compõem a ontologia. Um exemplo de uma ontologia fuzzy é apresentada na Figura 3.

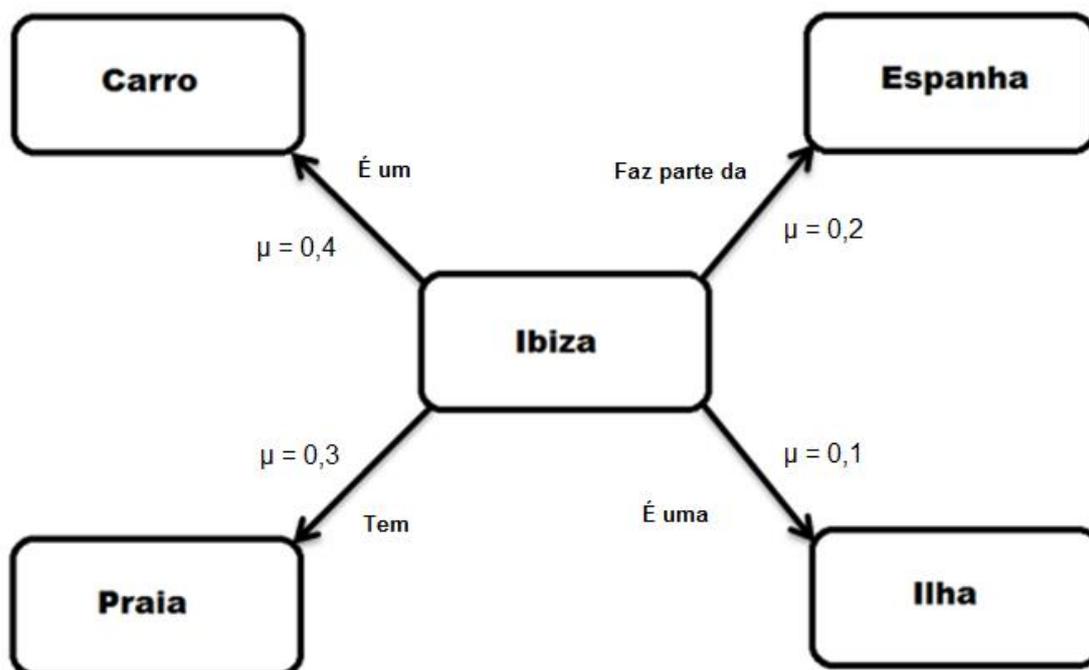


Figura 3: Exemplo de uma ontologia fuzzy. Imagem adaptada de Satler (2011, p.47).

Não é difícil perceber que a utilização da lógica fuzzy em uma ontologia pode dar a ela um grande poder de expressividade extra, uma vez que permite a resolução do problema do raciocínio impreciso nas ontologias. A ontologia demonstrada na figura 3 é o exemplo disso. Nela temos a visão exata do grau de pertinência de cada termo com relação à ideia central: Ibiza. Essa informação central possui relacionamentos com os termos que lhe atribuem significado, ou seja, quando falamos de Ibiza, podemos estar nos referindo a ilha, ou o carro, ou uma parte da Espanha, ou mesmo que ela possui uma praia. Esses termos, juntos, formam o grau 1 de pertinência.

Como as ontologias lidam diretamente com conceitos e termos utilizados na linguagem natural, seja de um idioma ou outro, é importante saber lidar com esse conhecimento. A análise dos termos presentes em documentos em linguagem natural, bem como sua correta manipulação na construção de uma ontologia, pode ser feita utilizando-se do chamado Processamento de Linguagem Natural, que define um conjunto de técnicas de análise de documentos escritos em linguagem natural.

O processamento de linguagem natural será o tema da próxima seção que mostrará suas técnicas e finalidades.

2.4. Processamento de Linguagem Natural

O processamento de linguagem natural, ou PLN, é tido como um conjunto de metodologias formais para análise de textos escritos em linguagem natural, ou seja, análise de textos em linguagem humana. A linguagem humana é essencialmente ambígua, com interpretações que dependem do contexto ou do ambiente no qual está inserido. Possui muitas regras gramaticais e conceitos abstratos que dificultam a análise por parte de sistemas computacionais em geral (SATLER, 2011, p. 88). Em outras palavras, pode-se dizer que PLN trata computacionalmente os aspectos da comunicação humana, os sons, palavras, sentenças e discursos, tudo isso considerando os formatos e as referências, as estruturas e os significados, os contextos e os variados usos da linguagem (GONZALEZ; LIMA, 2003).

Um dos objetivos do PLN é prover meios para os sistemas computacionais captarem dados resumidos dos textos em linguagem natural. Tal processo permite a redução de vocabulário dos textos em análise, gerando uma melhora acentuada na eficiência da recuperação da informação sobre os documentos. Contudo, esta redução de vocabulário pode acarretar na perda de informação sobre a representação dos documentos (SATLER, 2011, p. 88).

O PLN muitas vezes é confundido com a sua principal fase, o chamado pré-processamento. Também conhecido como normalização textual, o pré-processamento se encarrega de tarefas como: separação do texto em frases e posteriormente em palavras, conversão de símbolos e caracteres especiais, remoção de stopwords (palavras que não agregam significado ao texto, tais como artigos, preposições, conjunções, etc.), conversão de caixa (ALTA/baixa), dentre outras. A aplicação destas técnicas depende do tipo de análise a ser realizada no documento e demandam grande tempo de processamento, podendo chegar a 80% do tempo de processamento total. É nesta fase que todo o texto sofrerá análise, palavra por palavra, utilizando algumas ou todas as técnicas citadas anteriormente. A separação do pré-processamento do Processamento de Linguagem Natural não é muito clara. Algumas técnicas podem ser consideradas de pré-processamento por um autor, enquanto que outros podem discordar dessa consideração. O fato é que o pré-processamento é apenas uma parte do PLN.

Algumas tarefas realizadas no PLN são definidas a seguir (SILVA; CHAPETTA, 2003; ROMERO apud SATLER, 2011, p. 88):

- Análise léxica: nesta etapa ocorre a remoção de sinais de pontuação, acentos, etc.;
- Detecção de idioma: nesta etapa define-se qual o idioma predominante no documento;
- Verificação ortográfica: é realizada a análise dos termos, se estes seguem as regras gramaticais do idioma que foi detectado no texto;
- Redução de vocabulário: ocorre a remoção das stopwords e a extração da chamada raiz léxica dos termos (stemming);
- Sumarização de textos: permite a busca de atores para os modelos de casos de uso, bem como mostrar o conteúdo que pode ser considerado na descrição destes casos de uso.

As tarefas a serem executadas no PLN devem ser feitas numa ordem lógica, pois algumas possuem interdependência. Um esquema básico da sequência de aplicação das técnicas de PLN, é proposto na figura 4. Nela são descritos os passos básicos que podem ser seguidos para aplicar o PLN num documento.

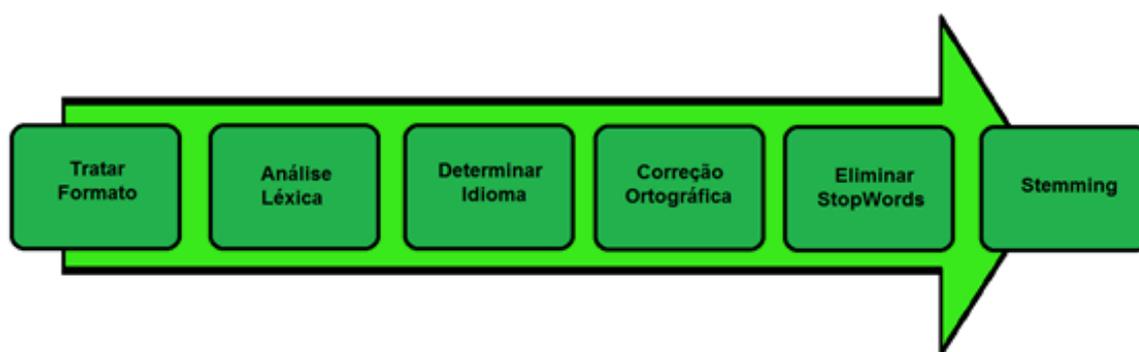


Figura 4: Esquema de aplicação das técnicas de PLN. Imagem adaptada de (SATLER, 2011, p.89)

A primeira etapa a ser executada é o tratamento dos formatos dos documentos que serão analisados. Geralmente ocorre alguma transformação para um formato padrão. O formato mais utilizado, nestes casos, é o “.txt”, por ser simples e admitir apenas informações textuais. Na segunda etapa ocorre a análise léxica, que fica

responsável por retirar acentos, dígitos, pontuação, etc. de cada palavra extraída do documento em questão.

A terceira etapa diz respeito à detecção de idioma, que pode ser realizado por diferentes técnicas. Romero (2008 apud SATLER 2011, p. 89) cita duas alternativas: a detecção realizada através de “trigramas” e através de “palavras comuns”. Ambas técnicas possuem valores de eficácia e de processamento parecidos.

Realizada a detecção do idioma, na quarta etapa ocorre a verificação ortográfica que pode ser feita através de duas tarefas: detecção de erros e correção automática destes erros. O objetivo desta etapa é buscar palavras com erros no texto e corrigi-las, trocando-as por palavras que mais se adequam ao contexto; estas palavras podem vir de dicionários padronizados para correção ortográfica.

Logo em seguida ocorre à eliminação de “stopwords”, o que pode reduzir o tamanho do documento de 30 a 50% ou mais. As stopwords, são palavras que não agregam significado semântico ao texto, tais como artigos, pronomes, etc., pois além de aparecerem muitas vezes durante um texto, servem de ligação entre verbos e outras frases e ideias gerais de um texto.

Por fim, ocorre a aplicação de “stemming” que, nada mais é do que a eliminação de prefixos e sufixos das palavras, ou seja, a redução das palavras à sua raiz léxica. A eliminação de stopwords e a aplicação do “stemming” têm como objetivo a redução do vocabulário do documento sem a perda de seu significado (SATLER, 2011, p. 90).

As técnicas e etapas aqui apresentadas podem ser realizadas ou não em sua totalidade, dependendo do tipo de tratamento pretendido para o texto, bem como a ordem de realização das mesmas pode variar.

A ideia central no PLN é a redução dos componentes a serem analisados, bem como prover técnicas que permitem interpretar a linguagem humana. A seleção das características mais importantes presentes em um texto diminui a quantidade de informação a ser processada e melhora a interpretação de seu significado. A análise de informações em linguagem humana por computadores fica mais fácil e se cria uma espécie de ponte que liga a linguagem humana à linguagem computacional. Porém, para se criar esta ponte entre estas duas áreas, é necessário o conhecimento tanto na área computacional, das técnicas de PLN, como na área linguística, do idioma presente nos documentos tratados.

Cada idioma tem suas peculiaridades e problemas diferentes que devem ser tratados de formas diferentes. Uma aplicação voltada para um determinado idioma, por exemplo o inglês, pode não ser adequada para o português. As duas línguas se diferenciam nas normas gramaticais, fonéticas, etc., em outras palavras, uma solução bastante plausível para um determinado idioma necessita de ajustes para tratar as peculiaridades de outro idioma. Cada língua tem seus problemas que dificultam de maneira diferente o seu correto processamento e interpretação.

2.4.1. Níveis de PLN

Num sentido amplo pode-se afirmar que o PLN tem o objetivo de fazer o computador se comunicar através da linguagem humana. Sendo assim, o PLN utilizado em determinado documento pode ser diferenciado em níveis que são definidos a seguir: (GONZALEZ; LIMA, 2003; LIDDY apud BEPPLER, 2008, pag. 33)

- Fonológico e fonético: relacionamento das palavras com os sons produzidos na sua pronúncia, os fonemas;
- Morfológico: análise das diferentes formas de uma palavra, classificação das palavras em categorias morfológicas;
- Sintático: a construção de sentenças através de frases e a análise do relacionamento entre as palavras, cada uma assumindo seu papel estrutural nas frases;
- Léxico: se relaciona com o significado puro da palavra e de sua estrutura;
- Semântico: a inter-relação entre as palavras e seus significados e a junção das mesmas para a geração dos significados das sentenças;
- Pragmático: o uso de frases e sentenças em contextos diferentes, o que afeta o significado, ou seja, aplicação de conhecimento externo tanto no documento como na consulta.

Sistemas de busca com linguagem natural, mesmo tendo a aplicação de todos os níveis de PLN citados anteriormente, possuem pouca clareza da informação podendo esta ainda estar implícita, necessitando de deduções. Há uma grande

dificuldade na determinação do que se deseja buscar, pois uma consulta pode apresentar pouquíssima informação, o que torna a busca, muitas vezes, subjetiva. Deste modo, raramente o PLN é usado em sistemas de recuperação de informação e sim como suplemento para modelos estatísticos (BEPPLER, 2008, pag. 33).

As técnicas de PLN possuem aplicação nas mais variadas áreas e permitem gerar bases interessantes para muitas análises diferentes, podendo ser integradas com ontologias para eliminação de redundâncias, padronização na saída dos dados, etc., além de possibilitar ao computador maior potencial de representação e abstração de dados e entendimento de seus significados, em textos e documentos diversos.

2.4.2. Frequência relativa de termos

Ainda dentro do contexto do PLN e segundo Wives (1999, p.23), a técnica mais utilizada para verificação das palavras ou atributos mais marcantes ou significativos em um texto é a frequência relativa. Esta técnica define, através do número de ocorrências de uma palavra em um texto, o quão relevante esta palavra é para o referido documento. Ou seja, o quanto esta palavra é importante na representação do significado do documento. A figura 5 apresenta a fórmula utilizada no cálculo da frequência relativa.

$$F_{rel}x = \frac{F_{abs}x}{N}$$

Equação 2: Fórmula da frequência relativa (WIVES, 1999, p.23)

Na figura 5 podemos observar que a frequência relativa de uma palavra x ($F_{rel}x$) em um documento é dada pela divisão da frequência absoluta da palavra x ($F_{abs}x$) pelo número total de palavras que ocorrem no texto (N).

Este cálculo geralmente é executado como parte do Processamento de Linguagem Natural, onde para cada termo extraído de um documento é calculado sua frequência relativa.

Para implementar as técnicas de PLN citadas nesta seção, é necessário utilizar uma linguagem robusta que possibilite a manipulação eficiente das tarefas a serem realizadas. A linguagem JAVA, por ser orientada à objetos, torna a criação da ontologia fuzzy mais simples, além de propiciar uma maneira mais intuitiva de representar os portfólios e seus conceitos.

Por esta razão, na próxima seção será abordada a linguagem de programação JAVA, citando suas peculiaridades e vantagens de uso para este trabalho.

2.5. Linguagem de Programação JAVA

A linguagem de programação JAVA é orientada a objetos e possui uma versatilidade surpreendente. Foi desenvolvida pela Sun Microsystems com uso inicial voltado para aplicações em aparelhos eletrônicos de consumo como fornos micro-ondas, agendas eletrônicas, etc. O objetivo de sua criação nasceu do desejo de uma linguagem mais simples e eficiente que as predecessoras, porém, apesar de conseguir cumprir estes requisitos, não se tornou popular comercialmente. Os criadores da linguagem se viram obrigados a adaptá-la para uso em microcomputadores conectados na internet. Assim Java permitiu uma revolução com a criação dos applets (programas com a capacidade de trocar dados na web utilizando-se da estrutura de um web browser) o que tornou a linguagem popular, atraindo milhões de desenvolvedores. Diversas áreas como a de banco de dados, passaram a buscar formas de integrar-se ao novo paradigma proposto na linguagem: aplicações voltadas para o uso de redes de computadores (INDRUSIAK, 1996).

Como linguagem orientada a objetos, JAVA é líder de mercado tanto em número de aplicações como em número de desenvolvedores (INE, 2016b). C++ é a segunda mais utilizada que suporta o paradigma de orientação a objetos, porém a linguagem Java é mais indicada por apresentar vários aspectos que não podem ser encontrados na linguagem C++. Segundo Almeida (2005), quatro desses aspectos são mais relevantes para o uso de Java em detrimento de C++:

- Independência de sistema operacional: característica também conhecida como portabilidade. Permite que um código em Java possa ser executado em máquinas diferentes e gere um mesmo resultado independente do sistema operacional utilizado por elas. Isso é possível devido aos chamados bytecodes, ou seja, a tradução do código de alto nível para código de baixo nível realizado pelo compilador de Java, o javac;
- Performance numérica: alguns testes realizados em condições idênticas para resolução de problemas específicos, mostrou que a linguagem Java é mais eficiente que a C++ para um problema análogo, as conclusões de outros autores que realizaram estudos parecidos é que ambas são equivalentes em eficiência;
- Capacidade de reutilização de software: os desenvolvedores Java estão focados na criação de novas classes e a reutilização de classes existentes. Diversas bibliotecas estão disponíveis com conjuntos de classes prontas para uso; o que torna o desenvolvimento em Java mais rápido e focado na resolução de novos problemas com a possibilidade de utilização de soluções já bem definidas e testadas sofrendo apenas adaptações;
- Suporte a persistência de dados: a linguagem Java tem API's (Application Program Interfaces) com a capacidade de criar, ler e editar arquivos XML, que são uma forma de armazenar permanentemente os dados em aplicações de forma portátil que são baseados em código aberto, tornando Java muito mais poderosa em uma vasta gama de aplicações.

A linguagem Java é muito versátil, podendo ser utilizada para a criação de diversos aplicativos de uso geral. É muito poderosa em ambientes distribuídos complexos como a internet. Muitos avanços na computação mundial foram impulsionados pela linguagem Java: acesso remoto a banco de dados, banco de dados distribuídos, interatividade em páginas web, gerência de documentos, network computer, EAD (ensino à distância), jogos e entretenimento, etc. (INDRUSIAK, 1996)

Alguns quesitos pesaram na escolha desta linguagem para implementação da ferramenta proposta nesse trabalho de conclusão de curso. Os principais são:

reutilização eficiente de código, simplicidade de codificação e portabilidade. Além disso, pesam na sua escolha a diversidade de recursos para desenvolvimento tais como as centenas de fóruns de discussão, as máquinas virtuais como *ECLIPSE*, entre outros.

Com relação à reutilização eficiente de código, citada anteriormente, mais uma vantagem no desenvolvimento em linguagem JAVA está na simplicidade na reutilização de padrões de projeto, que é o tema abordado na próxima seção.

2.6. Padrões de Projeto

Padrões de projeto são metodologias ou soluções destinadas à resolução de diversos tipos de problemas presentes no desenvolvimento de aplicações (MINETTO, 2007). São soluções que foram testadas e aprovadas na resolução de problemas específicos. Geralmente a estrutura de um padrão de projeto deve ser específica o suficiente para resolver um dado problema, mas com componentes genéricos que possam ser reutilizados para atender a problemas futuros.

Segundo Gamma e outros (1995 apud GONÇALVES, 2004 p. 37), utilizar padrões de projeto em sistemas orientados a objetos, os torna mais flexíveis e reutilizáveis. Eles auxiliam os projetistas a reutilizar técnicas, ideias e estruturas que deram certo em determinados projetos na construção de outros projetos.

Algumas vantagens podem ser obtidas com a utilização de padrões de projetos, tais como: aumento de produtividade, redução de complexidade dos sistemas, uniformidade de software, reuso de software, entre outras (GONÇALVES, 2004).

Pensando nestas vantagens, foi escolhido um padrão de projeto para o desenvolvimento da ferramenta, o padrão de projeto MVC. Este padrão será abordado na próxima subseção.

2.6.1. Padrão de Projeto MVC (*Model-View-Controller*)

É um padrão de projeto muito difundido por gerar uma separação lógica do desenvolvimento de aplicações em três camadas básicas: Modelo ou *Model*, Visão ou *View* e Controlador ou *Controller*.

Segundo Viana (2008), a funcionalidade básica de cada camada é a seguinte:

- A camada *Model* é responsável por armazenar as informações referentes ao domínio e os objetos que geram a funcionalidade principal dos sistemas de informação;
- A camada *View* é responsável pelos recursos de visualização, ou seja, os recursos e objetos que realizam a interface entre o sistema e o usuário;
- A camada *Controller* fica encarregada de realizar a mediação entre a camada *View* e a camada *Model*. Ela define a maneira como a camada *View* deve agir ao receber informações vindas do usuário, manipulando o estado dos objetos da camada *Model*.

A divisão de desenvolvimento em camadas permite que a criação do software possa ser feita por programadores trabalhando em separado. Um programador de banco de dados, por exemplo, pode gerar toda a lógica de acesso ao banco, atuando diretamente na camada *Model*, enquanto que, em paralelo, um programador especialista em interface gráfica, gera toda a comunicação do software com o usuário. Por fim, pode ser criada por um outro programador a camada *Controller* que vai gerir toda a comunicação entre as camadas com as regras para esta comunicação. Esta separação entre a parte lógica funcional e a apresentação é a principal vantagem na utilização deste padrão, além de proporcionar uma melhora no trabalho em equipe e agilidade no desenvolvimento (MINETTO, 2007).

Neste trabalho, o padrão MVC pode reduzir, consideravelmente, a complexidade no desenvolvimento da ferramenta, motivo pelo qual ele foi escolhido. Na próxima seção, serão abordadas algumas propostas encontradas na literatura que estão relacionadas à temática foco a ser desenvolvida neste trabalho.

2.7. Trabalhos Relacionados

Neste tópico serão abordados e discutidos dois trabalhos que estão relacionados com o presente trabalho de conclusão de curso. O primeiro intitulado como *Um estudo sobre agrupamento de documentos textuais em processamento de informações não estruturadas usando técnicas de “Clustering”* é uma dissertação de Wives (1999) para obtenção do título de Mestre em Ciência da Computação. O segundo intitulado *Enhancing Portfolio Assessment: An Application of Fuzzy Ontologies* é um artigo publicado por Satler e outros (2011).

2.7.1. Um estudo sobre agrupamento de documentos textuais em processamento de informações não estruturadas usando técnicas de “Clustering”

Neste trabalho Wives (1999) realizou um estudo sobre métodos de agrupamento de documentos no formato ASCII, que são organizados automaticamente em grupos de documentos similares como intuito de facilitar a localização, manipulação e análise por parte do usuário. Algo interessante que este trabalho apresenta é uma comparação entre algoritmos de agrupamento e a posterior apresentação de um algoritmo superior desenvolvido pelo autor que pode contribuir consideravelmente para o presente trabalho. Apresenta-se uma ferramenta de agrupamento de textos como um resultado final e uma interessante fórmula de cálculo de similaridade entre os objetos com uma abordagem fuzzy.

A proposta para agrupamento textual conta com etapas bem estabelecidas pelo autor e engloba áreas como o processamento de linguagem natural. Contudo, o autor reforça que não são realizadas etapas de pré-processamento linguístico. Ele simplesmente considera que os documentos foram preparados anteriormente com o auxílio de um corretor ortográfico.

Em linhas gerais, os problemas inerentes à linguagem utilizada nas buscas em geral, realizadas principalmente na internet, estão em uma classe de problemas denominados “problemas de vocabulário”. Esta classe de problemas está relacionada com a não existência de um vocabulário padrão para a publicação de documentos, o que dificulta a busca de informações através de palavras-chave. Buscando contornar este problema, Wives realizou um estudo dos documentos, retirando deles palavras

relevantes quanto ao seu conteúdo, categorizando-as através de características analisadas e selecionadas, e agrupando-as usando um cálculo de similaridade entre os objetos, ou seja, os termos presentes nos documentos.

2.7.2. Enhancing Portfolio Assessment: An Application of Fuzzy Ontologies

Satler e outros (2011) observaram que a abordagem de uso de Portfólios surgiu como uma importante ferramenta de auxílio no processo de aprendizagem. Contudo, os autores afirmam que as ferramentas de apoio à avaliação destes portfólios estão longe de atingir as expectativas. Tendo isso como problema central do artigo, eles propõem em seu trabalho uma estrutura baseada em ontologia fuzzy para o apoio à avaliação de portfólios eletrônicos.

O experimento realizado pelos autores consistiu na implementação de um aplicativo que recebia como entrada a representação de uma coleção de portfólios de alunos de um determinado curso e o material didático utilizado no curso. Posteriormente, esse material foi usado para gerar representações conceituais computacionais que seriam utilizadas para criar um relatório de avaliação para cada portfólio. Estes relatórios gerados servem como apoio na tarefa de avaliação individual de cada portfólio.

Para aplicar esta metodologia, foram gerados um conjunto de conceitos ponderados e uma ontologia fuzzy para o curso e para cada portfólio de cada aluno. O conjunto de conceitos, seus pesos e relacionamentos são representados pela ontologia fuzzy gerada, que caracteriza a representação conceitual dos artefatos a serem estudados.

Nos testes realizados, os autores afirmam que obtiveram um grau elevado de correlação da avaliação feita pelo professor (o julgamento humano) e a obtida através da ferramenta (a avaliação através de computadores).

O trabalho realizado pelos autores se aproxima mais das aspirações desejadas neste trabalho de graduação. Possui elementos de avaliação de portfólios munidos de uma ontologia fuzzy que, por ter um caráter flexível, auxilia na geração e utilização de regras para a avaliação.

Duas diferenças básicas podem ser citadas entre este trabalho de conclusão de curso e o trabalho de Satler e outros autores: A avaliação de portfólios de idiomas diferentes (neste trabalho o idioma dos documentos dos portfólios é o português, enquanto que o de Satler e outros autores o idioma utilizado foi o espanhol); e a linguagem utilizada para gerar a ferramenta (A linguagem utilizada neste trabalho é a JAVA e no outro, a linguagem utilizada foi a linguagem C++), além de outras peculiaridades e particularidades de aplicação e testes.

Apesar de parecerem sutis, estas diferenças trazem grandes modificações nos problemas tratados, principalmente em relação ao idioma, pois ele modifica as técnicas de PLN utilizadas e até mesmo a forma de utilização destas técnicas.

Na próxima seção serão discutidos o problema em questão neste trabalho e a proposta de solução para este problema.

3. Proposta

Nesta seção serão abordados e explicitados o problema foco deste trabalho bem como o modelo utilizado como proposta de solução. O objetivo principal é gerar uma descrição detalhada das partes que compõem o modelo conceitual e definir quais etapas e ferramentas serão utilizadas para gerar a solução final.

3.1. Definição do problema

De forma mais específica, o problema que este trabalho se dispõe a resolver é prover um meio de auxiliar professores na avaliação de portfólios digitais. O Meio encontrado foi geração de uma ferramenta de avaliação automática de portfólios digitais. Esta ferramenta pretende realizar a avaliação através de análises de documentos gerados eletronicamente, o que categoriza uma avaliação de e-portfólio. O porquê de se auxiliar o professor no processo avaliativo está no fato de que este processo é de uma complexidade singular. Ele demanda esforços para selecionar e rever documentos digitais, relatórios, protótipos, etc. de diversos alunos, avaliando de forma individual a maioria das evidências. Além disso, na literatura não há muito

suporte para a avaliação de portfólios. Desta maneira, o professor fica sobrecarregado, tendo que gerir todo o processo sozinho, sem nenhum suporte. A realização destas atividades não é uma tarefa trivial por necessitar de uma visão individual de cada aluno na verificação de certos objetivos a serem alcançados na aprendizagem. Além disso, é necessária uma visão geral da turma em si, buscando erros que possam ser corrigidos no processo de aprendizagem utilizado.

Muitos são os esforços para a utilização de portfólios como modelos pedagógicos. Cada disciplina tem peculiaridades, o que demanda adaptações no processo, dificultando e gerando uma necessidade de maior domínio do assunto por parte do professor que aplica este modelo aos alunos. Este trabalho tenta minimizar as dificuldades no processo avaliativo, ainda sendo necessário que o professor molde os portfólios de forma que estes possam alimentar a ferramenta que será proposta para a resolução deste problema.

A ferramenta proposta têm como finalidade auxiliar o professor no processo avaliativo, ou seja, não é o intuito desde trabalho propor uma forma de substituir a avaliação humana dos artefatos. Ao contrário, está sendo proposto é uma ferramenta de suporte na correção dos artefatos dos alunos, de forma a auxiliar o professor no processo de avaliação dos portfólios.

Na próxima seção será abordado o modelo conceitual proposto como solução para o problema aqui descrito.

3.2. Modelo conceitual

O modelo conceitual utilizado neste trabalho tem referência no trabalho de Satler e outros (2011), com a utilização de etapas e processos parecidos, mas com foco e objetivos diferentes que serão apresentados e explicitados neste tópico.

O objetivo principal deste trabalho é realizar uma análise comparativa entre os documentos de portfólio gerados pelos alunos e as referências por eles utilizadas para gerar tais documentos. Como resultado, o sistema gera um relatório para auxílio na avaliação do portfólio. Este relatório é gerado através de uma análise semântica dos documentos do portfólio de cada aluno e dos documentos de referência dados pelo professor na disciplina.

Para facilitar a implementação da ferramenta, o sistema foi dividido em duas etapas específicas ou processos distintos interdependentes: “processo construtivo do portfólio” e “processo de geração de relatório”, como pode ser visto na figura 6 e descrito em seguida.

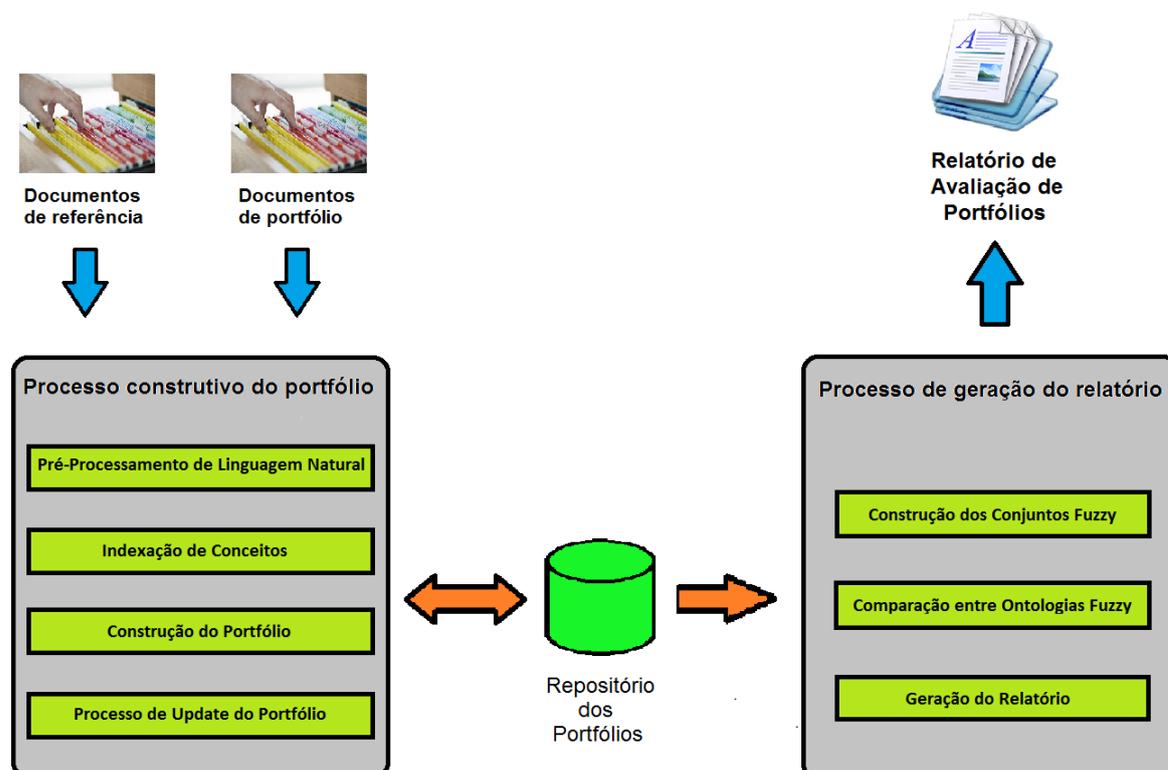


Figura 5: Representação do modelo de suporte para avaliação. Imagem gerada pelo autor.

Como pode ser verificado na figura 6, o sistema recebe como entrada os documentos de referência e os documentos desenvolvidos pelos alunos, que são usados na primeira etapa que é o processo construtivo do portfólio. Os resultados desse processo são versões conceituais dos portfólios de cada aluno e dos documentos de referência. Todos estes portfólios conceituais ficam armazenados no repositório de portfólios. Em seguida, o processo de geração do relatório compara os portfólios dos alunos com o portfólio de referência e gera como saída o relatório de avaliação dos portfólios, de acordo com os parâmetros e algoritmos desejados.

Em seguida será especificado em pormenores cada parte de cada processo do sistema de geração de construção do portfólio.

3.2.1. Processo de construção do portfólio

A primeira coisa a ser feita é definir uma estrutura que representará o conteúdo do portfólio. Assim é realizado o processo construtivo do portfólio que faz a leitura dos documentos de entrada na íntegra e retira deles termos que lhe atribuem significado. Posteriormente são padronizados e interpretados pelo segundo processo, que é a geração do relatório em si.

A criação do portfólio se dá através das seguintes etapas:

- Pré-processamento de linguagem natural: nesta etapa é realizada a extração do conteúdo textual dos documentos, a extração dos termos;
- Indexação de conceitos: nesta etapa ocorre a indexação dos termos extraídos no passo anterior e que possuem informações sobre os documentos dos quais eles foram retirados;
- Construção do portfólio: utiliza os conceitos indexados e lhes atribui pesos de acordo com a frequência de ocorrência nos documentos e outros parâmetros para gerar o portfólio;
- Atualização de portfólio: permite a inserção de novos conhecimentos em um portfólio através de novos documentos adicionados a ele.

Realizadas todas estas etapas, o processo de construção do portfólio é finalizado e fica armazenado no repositório de portfólios que alimentará a próxima etapa.

3.2.2. Processo de geração do relatório

O processo de geração do relatório tem como base o repositório de portfólios que foi gerado no processo anterior. Após receber os dados presentes neste repositório, o processo de geração do relatório realiza as etapas descritas a seguir:

- Fuzzyficação: Nesta etapa ocorre a padronização dos dados presentes no repositório de portfólios para lidar com a imprecisão das informações;

- Modelagem para ontologia difusa: nesta etapa ocorre a modelagem dos dados para criar as relações entre eles, necessárias para a inserção da ontologia;
- Geração do relatório: Nesta etapa ocorre a análise das relações criadas na ontologia fuzzy. São calculados os graus de compatibilidade entre os portfólios comparando seus termos. Posteriormente, é gerado um relatório como saída.

É interessante que a saída gerada nos relatórios seja configurável. Por essa razão, optou-se por permitir a comparação tanto dos documentos de referência com os documentos específicos de cada aluno como também a comparação entre a documentação de um aluno com outro aluno, o que facilita na identificação de possíveis cópias ou plágios. Essa configurabilidade de comparação também possibilita diferentes formas de análises de desempenho, sendo elas: individual de cada aluno, geral de uma turma e de aluno para aluno.

Na próxima subseção será abordada a ontologia fuzzy, explicando a sua aplicação no modelo conceitual.

3.2.3. Repositório dos Portfólios

O repositório de portfólios é formado por um conjunto de ontologias fuzzy, que representam a descrição conceitual de cada portfólio. A utilização de uma ontologia fuzzy permite a construção de uma representação semântica do portfólio do aluno de acordo com seu conteúdo textual, ou seja, uma representação através de termos ou conceitos, que atribuem significado para o conjunto de documentos do portfólio em questão. Um portfólio possui inúmeros documentos de um dado assunto ou contexto, sendo analisados como um único conjunto de conceitos que representam este portfólio. Para gerar esta representação, além dos conceitos é necessário o cálculo do peso que cada um destes conceitos possui na agregação de significado ao portfólio, e o cálculo do relacionamento entre estes conceitos, tidos como os graus de pertinência da lógica fuzzy.

Assim a utilização da ontologia fuzzy permite realizar a representação dos portfólios através de conceitos extraídos dos documentos pertencentes a estes portfólios. Posteriormente, são realizados os cálculos de relacionamento entre os conceitos e o peso destes conceitos sobre o portfólio. Estes cálculos viabilizam a comparação semântica destes portfólios através de algoritmos de compatibilidade entre ontologias.

Mais detalhes sobre os cálculos citados e outras configurações serão apresentados nas próximas subseções.

Na próxima subseção discutimos uma representação do portfólio que será usada como o resultado do processo de criação do portfólio que foi discutido na subseção 3.2.1 deste trabalho.

3.3. Estruturas do Modelo Conceitual

A primeira tarefa a ser realizada na construção do modelo é a definição da estrutura que representará o conteúdo do portfólio. Para gerar essa representação são considerados apenas os documentos com informações textuais, ou seja, somente os documentos como artigos e textos diversos são considerados em detrimento de gráficos, imagens, vídeos e outros artefatos coletados. Isto ocorre porque na abordagem que foi utilizada necessita de conceitos que podem dar significado aos portfólios dos alunos. Imagens, vídeos, etc. podem não dar o retorno esperado, pela dificuldade de extrair deles termos ou palavras que os definam para construir os conceitos. Desta forma, a análise semântica destas mídias é muito complexa e inviável para a proposta deste trabalho.

3.3.1. Representação Conceitual do Portfólio

Na definição da estrutura, a abordagem proposta neste trabalho como representação do portfólio foi definida como uma estrutura do tipo $P(\gamma, OF)$, onde P é a representação do portfólio em questão, γ é o conjunto de conceitos relevantes que representam o contexto ou domínio no qual o portfólio está inserido e OF representa

uma conceitualização da ontologia difusa sobre este contexto. A ontologia difusa citada é caracterizada por uma rede valorada de relações entre os conceitos mais relevantes no portfólio. Estas relações são representações das conexões entre os conceitos no domínio em estudo e que estão presentes nos documentos que compõem o portfólio analisado.

Esta estrutura, na forma com que foi proposta, permite uma visão única de cada portfólio. Eles são analisados de forma separada, sendo gerada uma representação para cada portfólio. Desta forma, há a possibilidade de avaliação individual do desempenho dos alunos analisando seus artefatos produzidos durante a disciplina como um todo. Na estrutura temos a denotação de P_s onde P é o portfólio e s é o estudante, desta forma podemos identificar a quem este portfólio pertence para a avaliação individual deste aluno. Nos próximos parágrafos serão introduzidos mais detalhes sobre os componentes da estrutura.

A letra grega gama γ é dada como uma coleção de conceitos representados individualmente por C_i . Cada conceito C_i tem um peso correspondente representado por W_i em um portfólio P . O peso deste conceito no portfólio representa a sua utilidade em definir o conteúdo deste portfólio. Em outras palavras, o quão bom este conceito é em representar o conteúdo do portfólio, ou o quanto ele contribui na composição do significado deste portfólio. A equação 1, a seguir, demonstra essa coleção de conceitos e pesos.

$$\gamma = \{(c1, w1), (c2, w2) \dots, (cn, wn)\}$$

Equação 3: Representação da Coleção de Tuplas de Conceito e Peso.

Para que um conceito seja considerado relevante na representação do significado de um portfólio, é necessário que ele ocorra com mais frequência do que outros conceitos nos documentos pertencentes a este portfólio. Além disso, este conceito deve ocorrer de forma ocasional em outros portfólios. Ou seja, um conceito para ser relevante em um portfólio, necessita não só de ocorrer muitas vezes nos documentos pertencentes a este portfólio, mas ele também não pode ocorrer muitas vezes em outros portfólios. Caso isso ocorra, ele seria categorizado como um conceito

genérico. Conceitos genéricos são palavras que ocorrem muitas vezes em documentos, mas que não são consideradas stopwords; não agregam muito valor semântico aos documentos e não os caracterizam. Desta forma, é necessário obter o chamado grau de relevância, ou peso de relevância de cada conceito sobre um portfólio. Onde o peso w_i^p de um dado conceito c_i dentro de um portfólio p , pode ser calculado pela equação 2, dada a seguir.

$$w_i^p = \sum_{j \in p} f_{ij} \times \left(1 + \frac{\text{docs}(c_i, p)}{|D|} \right) \times \text{Ln} \left(\frac{|P|}{P(c_i)} + 1 \right)$$

Equação 4: Fórmula do Cálculo do peso dos conceitos no portfólio.

Na equação 2, apresentada anteriormente, f_{ij} corresponde a frequência de ocorrência de um dado conceito c_i em um documento d_j . Este documento d_j pertence ao portfólio p . O numerador $\text{docs}(c_i, p)$ se refere ao número de documentos no portfólio p em que o conceito c_i ocorre. $|D|$ corresponde ao número total de documentos considerados no portfólio em questão. $|P|$ é o número de portfólios pertencentes ao contexto em questão, ou seja, a quantidade de portfólios dos alunos acrescido do portfólio de referência para a avaliação em dado contexto. $P(c_i)$ é denotado pelo número de portfólios que o conceito c_i tem um grau de pertinência positivo, ou seja, o número de portfólios em que o conceito c_i agrega valor ao seu significado.

Calculados os pesos, abre-se a possibilidade da identificação da distribuição de relevância de todos os conceitos presentes no portfólio. Para a geração desta distribuição, os pesos que foram obtidos passam por uma normalização com seus valores dentro do intervalo: $[0, 1]$.

Retornando a estrutura proposta como representação do portfólio, o componente *OF*, a ontologia difusa do portfólio, pode ser considerado um conjunto de grafos dirigidos, uma vez que ela é uma rede valorada das relações dos conceitos relevantes ao portfólio. Cada nó nestes grafos representa um conceito e as bordas denotariam que um conceito “está relacionado com” outro conceito interligado. O valor

ou grau desta relação quantifica a intensidade da associatividade entre os conceitos relacionados e pode ser calculado pela fórmula exposta na equação 3, dada a seguir.

$$RD(c_i, c_j) = \frac{\sum_{d \in D} occur(c_i, d) \otimes occur(c_j, d)}{\sum_{d \in D} occur(c_i, d)}$$

Equação 5: cálculo do grau de similaridade dos conceitos.

Na equação 3 pode ser observado o cálculo do grau de associatividade entre os conceitos c_i e o conceito c_j . A base do cálculo é o somatório de ocorrências do conceito c_i em um documento d ao mesmo tempo em que o conceito c_j ocorre neste mesmo documento. Assim pode ser verificada uma relação entre estes conceitos dentro do conjunto de documentos pertencentes ao portfólio de forma isolada para cada documento. No cálculo, a função $occur(c_i, d)$ refere-se à ocorrência do conceito c_i no documento d e o valor de ocorrência pode ser obtido através da equação 4, definida a seguir.

$$occur(c_i, d) = \frac{f_{i(d)}}{f_{max(d)}}$$

Equação 6: padronização do número de ocorrências de um conceito em um documento.

Na equação 4, temos que $f_{i(d)}$ corresponde a frequência de ocorrência do conceito c_i do documento d e $f_{max(d)}$ é o número de ocorrência do conceito mais frequente no documento d . A equação 4 é similar à equação que foi apresentada na figura 5 da subseção 2.4 sobre processamento de linguagem natural.

Retornando a equação 3, temos um operador de conjunto fuzzy (\otimes). Este operador é chamado de diferença limitada, ou *bounded difference* (BD), sendo considerado uma t-norma. Seu modo de operação pode ser verificado na equação 5, definida a seguir.

$$x \otimes y = \max(0, x + y - 1)$$

Equação 7: aplicação da t-norma BD.

Os valores x e y da equação 5 aplicados na equação 3, seriam os valores retornados pelas funções de ocorrência $occur(c_i, d)$ e $occur(c_j, d)$ respectivamente. A equação 5 retorna sempre o maior valor positivo entre 0 e ∞ , ou seja, desconsidera valores negativos.

Foram demonstradas anteriormente algumas das equações utilizadas para o cálculo de parâmetros. Equações como o cálculo dos graus de similaridade entre conceitos e os cálculos intermediários para encontrar seu valor. Estes cálculos auxiliam na posterior geração do relatório de avaliação, que será o tema da próxima subseção.

3.3.2. Representação Conceitual do Relatório

O foco deste tópico é a definição da estrutura para a construção do relatório de avaliação dos portfólios dos alunos. A metodologia e o processo de geração de relatórios serão tratados em pormenores em sessões subsequentes.

O relatório dispõe basicamente de informações sobre o conteúdo do portfólio avaliado e um estudo comparativo deste portfólio com o portfólio dado como referência para o curso ou disciplina em questão. Este estudo comparativo auxilia nas tarefas de avaliação.

Neste trabalho optamos por gerar uma representação específica para cada aluno, ou seja, o conteúdo dos portfólios é tratado de forma separada. Esta estratégia agrega mais valor ao trabalho, pois permite a construção de uma visão individual do progresso de cada aluno no curso ou tema ao qual o portfólio se refere. Nos próximos parágrafos serão descritos os elementos para a representação do relatório.

Na primeira etapa do processo, o sistema obtém a representação interna de todos os portfólios dos alunos de acordo com as definições informadas na subseção anterior, e os agrupa formando um conjunto de portfólios de aluno definidos pelo símbolo ϕ , que pode ser verificada na Equação 6.

$$\phi = \{P1, P2, P3, \dots, Pn\}$$

Equação 8: Conjunto de portfólios dos alunos.

Após a definição do conjunto de portfólios dos alunos, utilizamos a mesma estrutura de geração do portfólio P (γ, OF), para a criação da chamada “representação de curso” (P_c). Esta representação é construída com base no conteúdo do curso, que pode ser constituído de quaisquer documentos textuais, referências bibliográficas com informações e conhecimento sobre o referido curso ou tema analisado. Este material deve ser fornecido preferencialmente pelo responsável do curso, geralmente o professor, para que a avaliação seja feita de maneira correta, utilizando os documentos que os alunos utilizaram como referência para gerar seus artefatos.

Na segunda etapa, é realizada a comparação entre a representação do curso, ou o portfólio do curso P_c e o portfólio de cada aluno do conjunto analisado, denotado por ϕ . Para realizar a comparação foi utilizado um procedimento baseado em um processo de “correspondência” (matching) entre duas representações de ontologia, no nosso caso, dois portfólios. Nessa correspondência levamos em conta o conceito de *grau de compatibilidade (GC)*. O grau de compatibilidade entre duas representações de portfólio P_i e P_j denotados por $GC(P_i, P_j)$ avalia o quão semelhantes ou próximas as duas representações de portfólio são semanticamente. Neste contexto, quanto mais semelhantes forem o portfólio do aluno denotado por P_s e a representação do curso denotada por P_c , maior será a qualidade do conteúdo do portfólio do aluno. Porém vale ressaltar que isso depende da constituição do portfólio do curso, ou seja, quanto melhor for sua base de conteúdo, com documentos que lhe atribuam real significado, mais próxima da realidade será a avaliação do portfólio do aluno.

O cálculo do grau de compatibilidade consiste em uma comparação entre duas ontologias, um método de comparação que foi inspirado por um conjunto de métricas para a medição de similaridade entre ontologias, proposto por Maedche e Staab (2002, p. 251-263 apud SATLER et al., 2011). As medidas propostas são para comparação de duas ontologias. Como neste trabalho utilizamos uma ontologia fuzzy, foi considerado que cada ontologia seria uma coleção de conjuntos fuzzy de um dado portfólio. Esta lógica será explicada com mais detalhes nos próximos parágrafos.

Seja C o conjunto de conceitos que compõem uma dada ontologia que é representada por O . Cada conceito δ_i onde $\delta_i \in C$ é definido como um dos componentes do conjunto difuso de C e pode ser caracterizada por uma função de pertinência definida na Equação 7.

$$\mu_{\delta_i}: C \rightarrow [0,1], i. e, \delta_i = \sum_{j=1}^m \mu_{\delta_i}(c_j)/c_j$$

Equação 9: Função de Pertinência dos Conceitos.

Na equação, δ_i é o conjunto fuzzy que representa um dado conceito c_i . Ou seja, δ_i é um conjunto de m conceitos e μ_{δ_i} é o valor numérico que representa um grau de similaridade entre c_j e δ_i . Desta forma podemos dizer que: $\mu_{\delta_i}(c_j) = RD(c_i, c_j)$.

A obtenção do grau de compatibilidade (GC) é realizada utilizando o coeficiente de Jaccard, que será detalhado posteriormente, de acordo com o conjunto de classe teórica. O procedimento pode ser visto em detalhes no algoritmo que é definido em seguida.

Algoritmo 1: Algoritmo de grau de compatibilidade

```

 $O^c$  é a ontologia que correspondente ao portfólio do curso  $P^c$ 
 $O^s$  é a ontologia que correspondente ao portfólio do aluno  $P^s$ 
 $|O^c|$  e  $|O^s|$  correspondem ao tamanho de cada ontologia respectivamente
for  $c \in O^s$  do
    if  $c \in O^c$  then
         $\delta_{O^c}$  = corresponde ao conjunto fuzzy que representa o conceito  $c$  na
            ontologia  $O^c$ .
         $\delta_{O^s}$  = corresponde ao conjunto fuzzy que representa o conceito  $c$  na ontologia
             $O^s$ .
         $f = \text{jaccard}(\delta_{O^c}, \delta_{O^s})$ 
        { $f$  representa o grau de compatibilidade entre os dois conjuntos fuzzy}
    end if

```

```

 $\beta = A^m(f)$  { $\beta$  representa a similaridade entre conceitos comuns}
end for
GC =  $(|O^c \cap O^s| \otimes \beta) * \lambda / \min(|O^c|, |O^s|)$  { $\otimes$  é a t-norma e  $\lambda$  é o fator de
multiplicação}

```

No algoritmo 1 $jaccard(\delta_1, \delta_2)$ representa uma função de Jaccard entre dois conjuntos fuzzy e $A^m(f)$ denota a média aritmética dos graus de similaridade dos pares de conceitos em (f).

3.3.3. Coeficientes de similaridade

Os coeficientes de similaridade têm por objetivo a medição da proximidade entre duas instâncias, onde quanto mais parecidas ou próximas estas instâncias são, maior é o valor do coeficiente. Os valores destes coeficientes, geralmente ficam compreendidos no intervalo $[0, 1]$, onde o valor 1 significa total similaridade e 0 nenhuma similaridade entres os objetos analisados.

Entretanto, existem casos que se faz necessário verificar exatamente o oposto; ou seja, pode ser necessário medir o grau de dissimilaridade, que nada mais é do que o quanto dois objetos são diferentes entre si. Essa medida é, portanto, o valor complementar do grau de similaridade. Desta maneira se denotarmos o coeficiente de similaridade como CoS, a medida de dissimilaridade será o seu complementar $(1 - CoS)$ (MEYER, 2002). Assim, se somarmos o valor do coeficiente de similaridade e o valor do coeficiente de dissimilaridade obteremos o valor máximo do intervalo, ou seja, o valor 1.

O emprego de coeficientes de similaridade é comum em agrupamentos de indivíduos que são separados por classes. Em uma mesma classe encontram-se indivíduos com características em comum, geralmente muito semelhantes em determinado aspecto (MEYER, 2002). Outra aplicação foi demonstrada por Barros (2007), que afirmou que os coeficientes de similaridade são a maneira mais simples de medir a diversidade de espécies de animais entre pares de locais. Desta forma é

possível medir a similaridade entre ambientes diferentes, comparando as espécies de animais que habitam nelas.

A aplicação dos coeficientes de similaridade é vasto e um dos mais utilizados é o coeficiente de Jaccard, que será abordado a seguir.

3.3.4. Coeficiente de Jaccard

O índice ou coeficiente de Jaccard é dito ser qualitativo, uma vez que ele não considera a abundância das espécies, ou seja, ele não quantifica as populações presentes (PILLAR, 1996). Outras características marcantes deste índice são a sua simplicidade, objetividade e facilidade de entendimento e implementação. É um coeficiente de similaridade antigo, por volta de 1908 segundo Meyer (2002), mas que ainda hoje é um dos mais utilizados devido as características citadas anteriormente.

De forma sucinta, sua forma de operação é dada pela equação a seguir (PILLAR, 1996):

$$J_{\alpha\beta} = \frac{a}{a + b + c}$$

Equação 10: Fórmula de *Jaccard*

Onde α e β são os dois objetos analisados, “a” é o número de componentes que estão presentes tanto em α como β ao mesmo tempo, ou seja, a interseção dos elementos comuns entre eles; “b” corresponde aos componentes que ocorrem somente em β e “c” representa os componentes que ocorrem somente em α . Em outras palavras, $a + b + c$ é o mesmo que a união dos componentes de α e β .

E é exatamente por estas características e pela sua forma de operação simples, que ele foi escolhido neste trabalho. A forma com que ele vai ser utilizado, bem como variações propostas para resultados diferentes, serão abordados no tópico 5 que tratará dos resultados e discussões. A seguir será abordada a definição de capas.

3.4. Definição de capas

Neste tópico serão apresentadas, de forma detalhada as etapas integrantes do modelo conceitual, definido anteriormente, em dois processos diferentes: “processo de criação do portfólio” e “processo de geração do relatório” que possuem etapas bem definidas interdependentes entre si e suas saídas alimentam as demais etapas. O objetivo desta seção é discutir de forma aprofundada os dois processos definidos na seção 3.2 intitulada Modelo Conceitual. Focando a discussão nas etapas realizadas por estes processos, o que elas geram de saída e quais as técnicas utilizadas para sua execução.

O primeiro processo a ser realizado é o processo de criação do portfólio, aqui entendido como sendo todo o processo de representação dos portfólios, bem como o sistema de criação e padronização da ontologia difusa que consistem vários estágios de processamento de dados, estruturação e representação do portfólio. O processo de criação do portfólio será o tema a ser abordado a seguir.

3.4.1. Processo de criação do portfólio

Nesta etapa é realizada a leitura dos arquivos que compõem os portfólios dos alunos bem como os arquivos de documentação usados como referência na disciplina em questão, com o objetivo de gerar a conceitualização destes documentos através de termos. Para tanto é gerada uma lista de termos que representem os documentos do portfólio com seus respectivos pesos de relevância. Estes pesos caracterizam os documentos por serem palavras que agregam significado a eles, que lhes atribuem valor semântico. E por pertencerem a um determinado portfólio, estes documentos agregam significado a este portfólio através de seus termos. Assim o conjunto de conceitos serve como parâmetro para analisar o contexto ao qual o portfólio está inserido, categorizando-o. Uma série de etapas foram executadas para gerar a saída deste primeiro processo, que é uma lista de tuplas com os conceitos e seus respectivos pesos dentro de cada portfólio presente no sistema.

A técnica proposta para representação do portfólio foi dividida em quatro etapas, a exemplo do que foi dito na seção 3.2.1, que são as seguintes: “pré-

processamento de linguagem natural”, “indexação de conceitos”, “construção do portfólio” e “processo de update do portfólio”. O sistema foi dividido desta maneira porque o processo de construção dos portfólios demanda alguns estágios de processamento de dados, estruturação e representação do portfólio através de seus termos, como demonstrado na subseção 3.2.1. A seguir será abordada a etapa de pré-processamento de linguagem natural, que é a primeira etapa a ser executada no processo de construção do portfólio.

I. Pré-processamento de linguagem natural

Nesta etapa o objetivo é extrair o conteúdo textual do portfólio e armazená-lo em forma de conceitos individuais. Para executar esta tarefa, é necessário primeiramente, lidar com os diversos tipos de formatos dos documentos (.docx, .pdf, .odt, .txt, etc.), transformando-os em algum padrão para que o programa possa interpretar.

Como é necessária apenas a extração do conteúdo do texto, ou seja, as suas palavras para que se tornem conceitos, não importando a forma como foi formatado o texto, neste trabalho, adotou-se o padrão “.txt” devido a sua simplicidade em tratar o texto e por aceitar apenas informações textuais, sem gráficos ou imagens. Assim os documentos dos portfólios são convertidos de qualquer extensão (.docx, .pdf, etc.) para o padrão “.txt” e posteriormente é executado o pré-processamento.

Definido o padrão de formato de documentos a ser utilizado, dá-se início ao pré-processamento de linguagem natural. A primeira tarefa é a abertura do canal de leitura dos documentos. Aberto este canal, a próxima tarefa é a remoção das informações não textuais dos documentos lidos, tais como dígitos, datas, pontuação, etc. Esta etapa é chamada de “análise léxica”, que consiste na análise das linhas de caracteres dos documentos gerando os chamados “tokens”, que poderão ser manipulados mais facilmente pelas outras tarefas a serem realizadas.

Logo em seguida são utilizadas técnicas para redução do vocabulário, o que auxilia na representação de textos tornando-os mais significativos. Uma destas técnicas é a “remoção de stop words” (palavras que não agregam muito significado ao texto por serem palavras muito genéricas, utilizadas apenas para ligar um conceito a outro, tais como: artigos, pronomes, verbos de ligação, etc.). Desta forma, o conjunto

de conceitos fica mais compacto e com mais possibilidade de gerar o valor adequado de significado ao portfólio analisado.

Outras técnicas podem ser realizadas durante esta etapa, tais como: correção ortográfica, stemming e detecção de idioma, mas neste trabalho não são necessárias, uma vez que os textos a serem analisados já estão em português e a correção ortográfica não será aplicada para gerar um resultado mais real. Na próxima seção será abordada a indexação dos conceitos obtidos nesta etapa.

II. Indexação de conceitos

Nesta fase o objetivo é fornecer uma estrutura de índice contendo informações sobre todos os conceitos que foram gerados na fase anterior. A estrutura gerada é composta por um conjunto de objetos do tipo “termo”, ou seja, cada palavra obtida na fase anterior será interpretada como um “termo”. Cada objeto do tipo “termo” tem um ID (neste caso o próprio conceito) e uma lista de elementos que contêm informações relacionadas aos documentos nos quais estes conceitos estão presentes. As informações armazenadas nos elementos são: *identificação do documento* e o *número de ocorrência do conceito mais frequente neste documento*. Desta maneira temos informações em quais documentos o conceito aparece, bem como qual o conceito mais frequente em cada documento. Isto permite o cálculo padronizado das ocorrências do conceito em um documento, como ilustrado na equação 6, e posteriormente o cálculo do grau de similaridade, mostrado na equação 5.

A indexação dos conceitos ocorre com a absorção do conteúdo gerado na fase anterior, pré-processamento de linguagem natural, que gerou uma lista de conceitos com sua respectiva frequência de ocorrência em cada documento analisado. Este fato possibilita a obtenção das informações que nutrem os elementos citados anteriormente, o que viabiliza a indexação de conceitos de acordo com o que foi exposto e viabiliza o próximo passo, a construção do portfólio, que será tratado a seguir.

III. Construção do portfólio

Nesta etapa, são utilizadas as estruturas de índice dos conceitos geradas na fase anterior como fonte de dados para gerar os seguintes componentes: “conjunto de conceitos e pesos” simbolizados por γ , ou de forma separada denotados por c_i e w_i , respectivamente. Nesta fase também são calculados os graus de similaridade entre os conceitos, ou seja, o grau de relacionamento entre os conceitos do portfólio.

Para gerar o componente γ , os pesos de cada conceito obtido nos documentos do portfólio são calculados utilizando a equação 4, ilustrada na subseção 3.3.3.. Esta equação retorna o valor do peso que cada conceito possui dentro do portfólio analisado, possibilitando a construção do conjunto de tuplas conceito-peso do componente γ .

Cada conceito possui pesos diferentes quando analisamos portfólios diferentes, uma vez que estes portfólios são constituídos de documentos diferentes. Os conceitos, de um modo geral, dão significado ao portfólio, colocando-o dentro de um contexto. Desta forma, eles têm pesos de agregação de significado diferentes em cada portfólio. Um conceito c_i pode ser importante para compor o significado de um portfólio p_1 , mas pode ser pouco significativo para o portfólio p_2 , ou mesmo nem ocorrer neste portfólio.

Com o conjunto γ devidamente construído, o componente OF é gerado na forma de uma rede valorada, onde a ontologia difusa possui uma descrição das associações de grau de similaridade entre os conceitos, o RD (calculado através da equação 3 demonstrada na seção 3.3.3). Assim podemos obter a representação final do portfólio como também foi definido na seção 3.2.4: $P(\gamma, OF)$.

Construída a representação do portfólio, podemos ainda fazer uso de um mecanismo para atualizar este portfólio, de forma a adicionar conhecimento ao portfólio, mesmo que este já tenha sido analisado pela ferramenta. Este mecanismo de atualização é tratado de forma mais específica a seguir.

IV. Processo de atualização do portfólio

Este procedimento torna a estrutura proposta neste trabalho mais dinâmica, por permitir que novos conhecimentos possam ser adicionados na representação do portfólio através de novos documentos inseridos nele. Este processo pode ser realizado em todos os portfólios gerados na estrutura.

A primeira coisa a ser feita é o pré-processamento de linguagem natural e logo em seguida a indexação de novos conceitos para novos documentos. Após isso, alguns parâmetros são recalculados e atualiza-se a lista do componente γ com a inserção dos conceitos novos e a atualização dos pesos dos conceitos relevantes ao portfólio. Ao final, o componente OF é recalculado, ou seja, o conjunto de graus de similaridade dos conceitos.

Como pôde ser visto, este processo é simples, consistindo basicamente da atualização dos componentes gerados pelas outras etapas, com agregação de novos significados ao portfólio no qual os novos documentos foram inseridos.

Na próxima seção será abordado o processo de construção do relatório de avaliação de forma abrangente e detalhada.

3.4.2. Processo de Geração do Relatório

Depois de gerada toda a estrutura de representação dos portfólios, com seus conceitos relevantes e respectivos pesos, entramos na etapa onde é gerado o resultado final esperado, ou seja, o relatório de avaliação que guiará o professor no processo avaliativo do portfólio do aluno. Nesta subseção apresentamos a estrutura e o processo de construção do relatório de avaliação dos portfólios. Neste relatório serão encontradas informações sobre o conteúdo dos portfólios e um estudo comparativo entre as referências do conteúdo ministrado com o portfólio do aluno analisado, visando ajudar no processo de avaliação de portfólio, objetivo principal deste trabalho.

Este processo foi dividido em três partes a exemplo do explicitado na seção 3.2 que tratava sobre o modelo proposto para resolução do problema. As partes são: *Construção dos Conjuntos Fuzzy*, *Comparação entre ontologias fuzzy* e por fim a *Estrutura do Relatório de Avaliação*. Nas próximas seções serão abordadas cada parte com o detalhamento do seu uso.

I. Construção dos Conjuntos Fuzzy

O relacionamento entre os conceitos é calculado através da fórmula 3 definida na seção 3.3 Estruturas do Modelo conceitual. Este valor de relacionamento é armazenado numa estrutura chamada conjunto fuzzy que representa o relacionamento que um determinado conceito possui com cada um dos conceitos pertencentes ao portfólio. Assim ao final do processo de construção dos conjuntos fuzzy, tem-se um conjunto fuzzy para cada conceito presente no portfólio. De posse destas informações, é possível gerar a ontologia fuzzy, que consistem em um conjunto com todos os relacionamentos entre todos os pares de conceitos.

O valor de um relacionamento, ou RD, fica compreendido no intervalo $[0,1]$, levando em consideração em seu cálculo a ocorrência dos conceitos em cada documento pertencente ao portfólio analisado. Os relacionamentos cujo valor RD é igual a zero são desconsiderados, tanto nos conjuntos fuzzy quanto na ontologia fuzzy.

A geração dos conjuntos fuzzy e da ontologia fuzzy é repetida para todos os portfólios existentes, para que posteriormente possam ser comparados aos pares, como explicitado a seguir.

II. Comparação entre Ontologias Fuzzy

Depois de realizados os calculados e executados os passos para a valoração dos relacionamentos entre os conceitos de todos os portfólios, a próxima etapa é realizar a comparação da ontologia fuzzy pertencente ao portfólio de referência no curso com a ontologia fuzzy pertencente ao portfólio do aluno.

A comparação entre as ontologias ocorre de acordo com o algoritmo 1: definido na seção 3.3.1. Neste algoritmo, os vetores de conceitos são percorridos e para cada conceito que ocorra em ambos os portfólios é calculado o grau de compatibilidade entre os conjuntos fuzzy destes conceitos através do coeficiente de similaridade de Jaccard. O coeficiente analisa cada relacionamento em comum dos conjuntos fuzzy e gera o valor do grau de compatibilidade que fica compreendido no intervalo $[0,1]$. Posteriormente é feita uma média aritmética de todos os graus de compatibilidade encontrados e este valor é utilizado na fórmula do grau de compatibilidade dos portfólios.

O grau de compatibilidade entre os portfólios é o valor de similaridade dos portfólios analisados. Desta forma, quanto mais próximo ele esteja do valor máximo (100) mais semelhante estes portfólios serão. Este valor é uma das métricas que podem ser usadas para avaliar o aluno, mas não deve ser utilizado como nota final. É importante aqui considerar também a revisão do professor, servindo a métrica apenas como parâmetro de correção.

A seguir será abordada a estrutura do relatório de avaliação, que apresenta além do grau de compatibilidade entre os portfólios, outras informações acerca dos componentes presentes nos portfólios analisados.

III. Estrutura do Relatório de Avaliação

O Relatório de Avaliação dos Portfólios, basicamente contém as informações que irão apoiar a pessoa responsável pelo curso, tradicionalmente o professor, na tarefa de avaliar os artefatos gerados pelos alunos para a verificação do progresso na aprendizagem. O relatório dispõe de informações como: os conceitos mais relevantes utilizados pelo aluno na construção dos documentos do portfólio, o grau de similaridade entre estes conceitos e o grau de compatibilidade entre os portfólios analisados.

O relatório é dividido em três partes. Na primeira parte temos o cabeçalho que dispõem de informações gerais sobre o curso tais como: a data e horário de geração do relatório, o nome do aluno ao qual o portfólio avaliado pertence e nome do assunto analisado. Já na segunda parte, uma lista de conceitos mais significativos utilizados pelo aluno é apresentada, juntamente com uma lista de conceitos mais importantes referentes ao curso, o que fornece um primeiro parâmetro de comparação auxiliando no processo de avaliação. Estas listas estão organizadas em ordem decrescente, de acordo com o peso de relevância sobre o portfólio do aluno (Ps) e o portfólio de representação do curso (Pc).

Por fim é apresentado o grau de compatibilidade (GC) entre os portfólios analisados. Uma comparação das ontologias difusas que representam o portfólio do aluno e o portfólio do curso, que é um parâmetro importante para ajudar na avaliação de portfólio. Contudo, vale ressaltar que o grau de compatibilidade fornece apenas uma medida que deve ser usada como indicativo para a avaliação, mas não como

nota final do aluno na matéria ou curso, por ser uma avaliação do conteúdo textual bruto produzido pelo aluno, não realizando análises da construção dos documentos, ou seja, da forma com que foram apresentadas as ideias, além de não avaliar artefatos como imagens, vídeos, a participação do aluno nas discussões em sala, etc. Essas análises só podem ser realizadas pelo professor por serem de cunho subjetivo.

Na próxima seção será abordada a implementação da ferramenta proposta: um software gerado em JAVA como definido na subseção 2.5. Serão dados detalhes da implementação e decisões tomadas para gerar a estrutura proposta anteriormente.

4. Implementação

Nessa seção será apresentada a ferramenta para avaliação de portfólios que foi gerada utilizando o padrão de projeto MVC, descrito na seção 2.6.1 deste trabalho. Serão abordados detalhes de decisões de projeto e será apresentada a estrutura do software gerado com suas classes integrantes.

O padrão de projeto MVC gera uma separação lógica do desenvolvimento de softwares em três camadas básicas. Essas camadas são *model*, *view* e *control*. O objetivo deste padrão é proporcionar uma maneira de criar sistemas complexos dividindo-os em partes bem definidas que podem ser realizadas de forma independente. As camadas *model* e *view* podem ser construídas de forma totalmente separada e independente, uma vez que a camada *model* fica responsável pela criação da lógica de armazenamento das informações e as regras de negócio, enquanto que a camada *view* fica responsável pelos recursos de visualização, ou seja, a interface entre o sistema e o usuário. A comunicação entre estas duas camadas é realizada pela camada *control*, que por este motivo é criada após obter as informações da estrutura seguida pelas demais camadas.

Através desta estrutura, com este nível de separação lógica das partes integrantes do sistema, que a ferramenta de avaliação de portfólios foi constituída. Desta forma, a implementação será descrita ressaltando quais classes pertencem a cada camada e como elas se interagem.

Em linhas gerais o funcionamento do sistema começa com a criação lógica de uma estrutura que representa um “Portfólio”. Esta estrutura recebe os documentos e

extraí deles as palavras que são interpretadas como “termos”. Cada “termo” recebe além da palavra extraída do texto, informações dos documentos nos quais esse “termo” ocorreu e quantas vezes ele se repetiu. A extração das palavras é feita através de uma classe que realiza toda a lógica de comunicação com os arquivos em formato “txt”. Após a extração das palavras e contabilizada a sua ocorrência no documento, é realizada a remoção das chamadas “stopwords”, que são palavras que não agregam valor semântico aos documentos, ou seja, verbos de ligação, pronomes, artigos, etc. O resultado da remoção de “stopwords” é uma lista de “termos” que representam todos os documentos no portfólio analisado. Posteriormente são calculados os “pesos” que cada termo possui na constituição do significado do portfólio e são gerados os chamados “conceitos”, que nada mais são do que um “termo”, seu “peso” de representatividade neste “Portfólio” e uma estrutura que comportará o conjunto fuzzy, que será explicada a seguir.

Gerados os “conceitos” e calculados seus pesos no “Portfólio”, a próxima etapa é o calcular o “grau de similaridade”, ou simplesmente “RD”, entre os conceitos relacionados no portfólio. O “RD” é calculado para cada par de conceitos de um portfólio e leva em consideração a ocorrência dos conceitos em cada documento. O “RD” e as informações de identificação dos conceitos ao qual o “RD” pertence, ficam armazenadas nos “conjuntos fuzzy” dos objetos que representam cada conceito.

O processo de cálculo do “RD” se repete para todos os pares de “conceitos” pertencentes ao portfólio e ao final, cada conceito possuirá um “conjunto fuzzy” com informações sobre seu relacionamento com os demais conceitos. Em paralelo à geração dos conjuntos fuzzy é criada a “ontologia fuzzy” onde são armazenados todos os relacionamentos distintos dos pares de conceitos, ou seja, a união dos relacionamentos de todos os “conjuntos fuzzy”. As informações contidas na “ontologia fuzzy” são utilizadas para comparar portfólios diferentes gerando o chamado “*Grau de Compatibilidade*” que é calculado levando em consideração a ontologia de cada portfólio e a média aritmética do “grau de compatibilidade” dos conceitos que os portfólios analisados têm em comum. Este “grau de compatibilidade” entre os conceitos é calculado através do coeficiente de similaridade de “Jaccard”.

As próximas seções serão destinadas para explicar em detalhes como foi criada a ferramenta, quais classes e atributos foram utilizadas em cada camada. A

primeira camada a ser referenciada é a camada *model* e serão apresentados os diagramas que representam as classes integrantes desta camada.

4.1. Camada Model

Nesta camada são inseridas informações referentes ao domínio e os objetos que geram a funcionalidade principal do sistema com as regras de negócio do software. A seguir serão apresentadas as classes que foram criadas, neste caso as classes que lidam diretamente com os dados extraídos dos documentos que representam os portfólios. As informações brutas extraídas dos arquivos são armazenadas em uma classe denominada “Termo”; esta classe tem como atributos uma variável do tipo “String” que guarda cada palavra utilizada nos documentos do portfólio e uma lista do tipo “Lista” para armazenar a quantidade de ocorrência deste termo e o id do documento de referência, como pode ser visto no diagrama de classe a seguir.

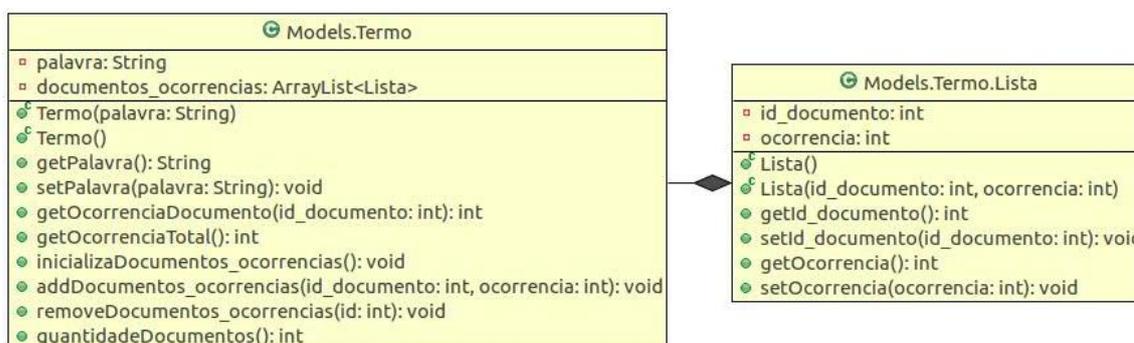


Figura 6: Diagrama de Classe – Classe Termo. Imagem gerada pelo autor.

A figura 7 apresenta a classe *Termo* com seus métodos e os seguintes atributos: uma variável do tipo *string* chamada *palavra* e um vetor de objetos do tipo “Lista”.

A classe *Lista*, armazena o id do documento no qual o termo aparece e a ocorrência deste termo no referido documento. O número identificador do documento

(id) tem referência na classe “Documento”, que possui basicamente o valor de ocorrência do termo que mais vezes se repetiu neste documento, o nome do documento que foi adicionado e seu id.

As informações extraídas e armazenadas na classe “Termo”, servem de parâmetros para o cálculo do peso de representatividade que o termo possui no portfólio e posteriormente a criação dos conceitos como pode ser visto na figura a seguir.

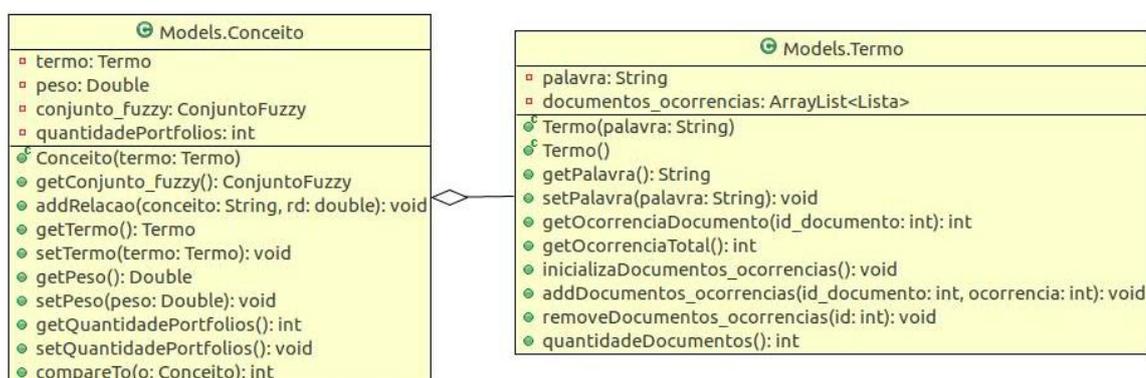


Figura 7: Diagrama de Classe – Classe Conceito. Imagem gerada pelo autor.

A figura 8 apresenta a classe “Conceito” que é composta por um objeto do tipo “Termo”, uma variável chamada *peso* que armazena o valor de representatividade do conceito no portfólio. Além disso, a classe “Conceito” possui um objeto chamado “conjunto fuzzy” que pode ser visto a seguir.

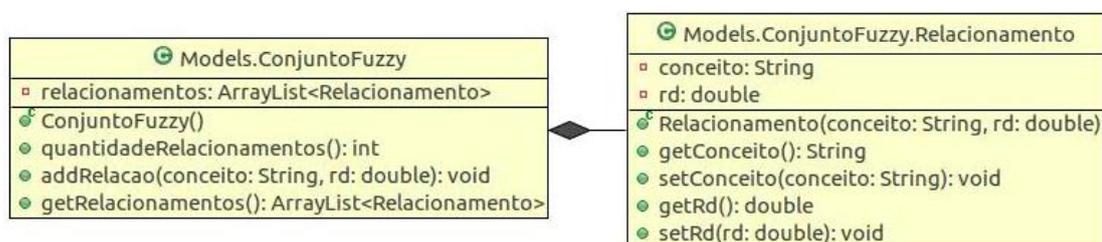


Figura 8: Diagrama de Classe – Classe Conjunto Fuzzy. Imagem gerada pelo autor.

O “Conjunto Fuzzy” armazena um vetor de relacionamentos. Cada objeto do tipo “Relacionamento” armazena o grau de similaridade entre o conceito ao qual o “Conjunto Fuzzy” pertence e o identificador do segundo conceito relacionado que é

representado pela “String” “conceito”. As informações dos relacionamentos são posteriormente utilizadas para calcular o “Grau de Compatibilidade” dos portfólios através das ontologias fuzzy, como foi explicado anteriormente.

A classe “Ontologia Fuzzy” representa a união (U) dos relacionamentos de todos os conjuntos fuzzy, ou seja, ela armazena todos os relacionamentos entre os conceitos, desconsiderando os relacionamentos repetidos. Essa classe pode ser vista com mais detalhes na imagem a seguir.

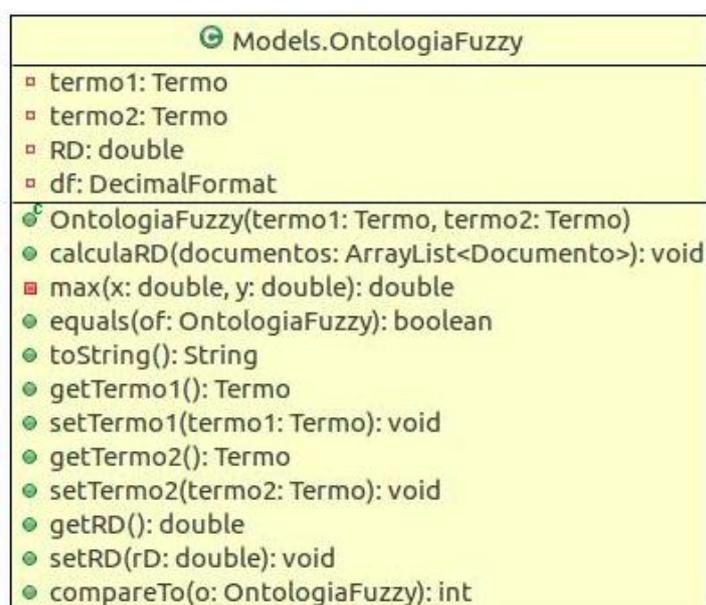


Figura 9: Diagrama de Classe – Classe Ontologia Fuzzy. Imagem gerada pelo autor.

A figura 10 apresenta a classe “Ontologia Fuzzy” que basicamente possui dois objetos do tipo “Termo” e o valor do grau de similaridade entre estes termos. Essa classe é parte integrante da classe “Portfólio” e dispõem de informações para o cálculo do “grau de compatibilidade” dos portfólios. A figura a seguir mostra a classe “Portfolio”.

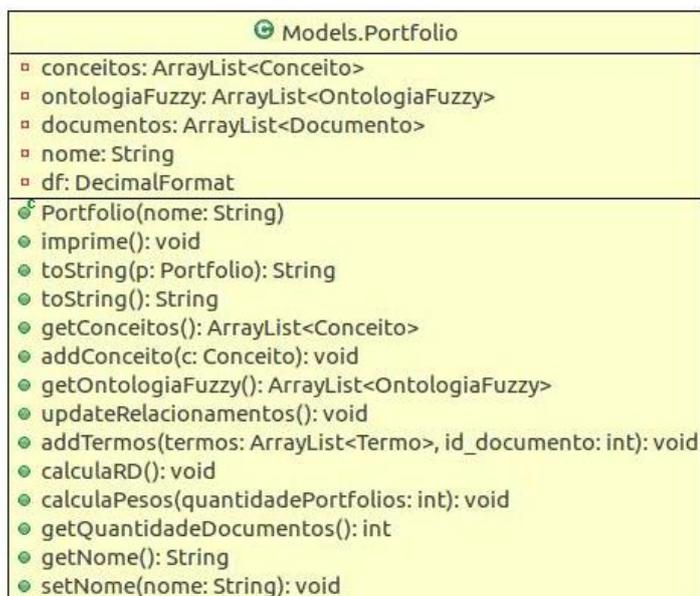


Figura 10: Diagrama de Classe – Classe Portfolio. Imagem gerada pelo autor.

A figura 11 apresenta a classe “Portfolio” que armazena todas as informações coletadas para construir o portfólio digital. Essa classe contém informações dos “Documentos”, os “Conceitos” extraídos destes documentos e a “Ontologia Fuzzy” que dispõem de todos os relacionamentos entre os conceitos deste portfólio. Ela possui métodos para calcular os pesos dos conceitos no portfólio, métodos para adição de novos conhecimentos neste portfólio (através de novos conceitos extraídos de novos documentos), bem como um método para realizar o cálculo do grau de similaridade entre os conceitos. Cada objeto do tipo “Portfólio” dispõem de basicamente todas as informações que um portfólio digital possui, o que facilita a criação do relatório de avaliação dos portfólios. Esse relatório é representado pela classe que pode ser vista a seguir.

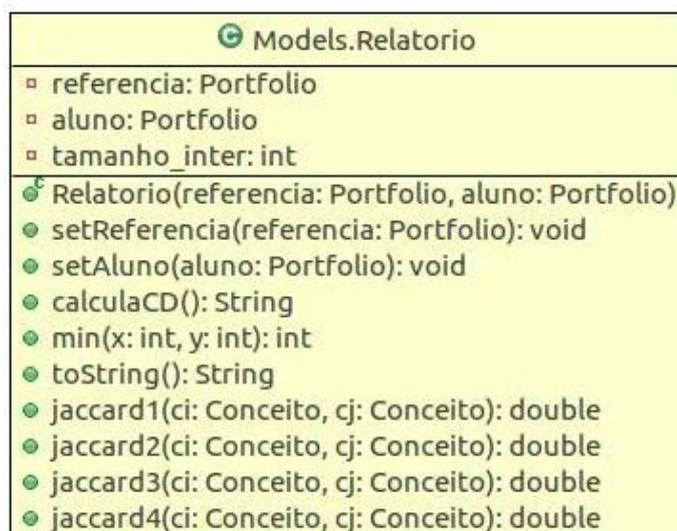


Figura 11: Diagrama de Classe – Classe Relatório. Imagem gerada pelo autor.

Na figura 12 podemos observar como foi constituída a classe Relatório. Basicamente essa classe recebe como atributos dois portfólios, realizando a comparação entre suas ontologias fuzzy e posteriormente gerando o grau de compatibilidade entre estes portfólios. Nela existem quatro métodos diferentes de Jaccard que retornam um valor *double* que corresponde ao grau de similaridade dos conceitos *ci* e *cj* que são passados como parâmetro. O valor de retorno destes métodos interfere diretamente no valor do grau de compatibilidade dos portfólios. Assim, para fins de teste, são gerados quatro valores de grau de compatibilidade dos portfólios analisados respectivamente.

Estas são as classes principais na camada *Model* que armazenam e geram as principais informações da ferramenta. Na próxima seção será abordada a Camada *View*.

4.2. Camada View

Nesta camada são criadas as classes que lidam diretamente com o usuário, ou seja, realizam a interface do usuário com o sistema. Elas fazem requisições para a camada *Controller* com o intuito de executar as operações solicitadas pelo usuário.

Todas as classes da camada *View* “conhecem” e interagem apenas com a camada *Controller*, acessando e utilizando seus atributos e métodos de forma direta,

mas não têm acesso direto aos atributos e métodos da camada *Model*. Estes, por sua vez, são acessados de forma indireta através da camada *Controller*.

A estrutura da camada *View* é composta por classes com atributos que juntos geram a parte visual do software. Estes atributos são objetos que dispõem de métodos especiais que ficam aguardando a interação do usuário para executar ações específicas definidas pelo programador.

A figura a seguir apresenta a “Janela Principal” do sistema que é constituída de uma classe.



Figura 12: Janela Principal da Ferramenta de Avaliação de Portfólio. Imagem gerada pelo autor.

Como pôde ser visto na figura 13, a janela do sistema possui menus que dão acesso as funcionalidades do sistema e ficam aguardando pela interação do usuário para executar as ações de acesso aos atributos e métodos da camada *Controller*. A figura a seguir mostra a classe que gerou a janela mostrada na figura 13.

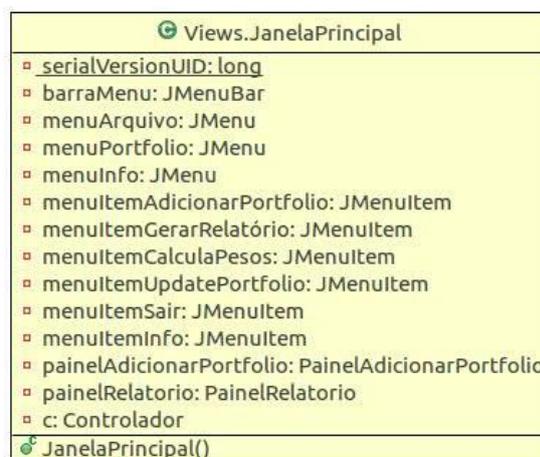


Figura 13: Diagrama de Classe – Classe Janela Principal. Imagem gerada pelo autor.

Os atributos dessa classe são basicamente os menus e submenus que estão presentes na janela principal do sistema onde o usuário pode solicitar qualquer funcionalidade.

Na próxima seção será abordada a camada *Controller* e sua comunicação com as camadas *View* e *Model*.

4.3. Camada Controller

Esta camada é responsável pelo controle da comunicação entre as camadas *View* e *Model* e armazena e gerencia todos os objetos, atributos e métodos gerados durante a execução da ferramenta. Essa classe possui apenas uma instância, ou seja, somente um objeto do tipo “Controlador” é criado na execução da ferramenta. As classes da camada *View* que necessitem de alguma informação ou precisem invocar algum método da classe “Controlador” acessam a mesma instância desta classe, ou seja o mesmo objeto e portanto, terão acesso as mesmas informações e não cópias, o que propicia uma redução da quantidade de memória gasta na execução das tarefas e garante que qualquer objeto acessará informações corretas, além de proporcionar uma melhor organização e controle dos dados e informações.

Na figura a seguir é possível observar a classe “Controlador” com seus atributos e métodos.

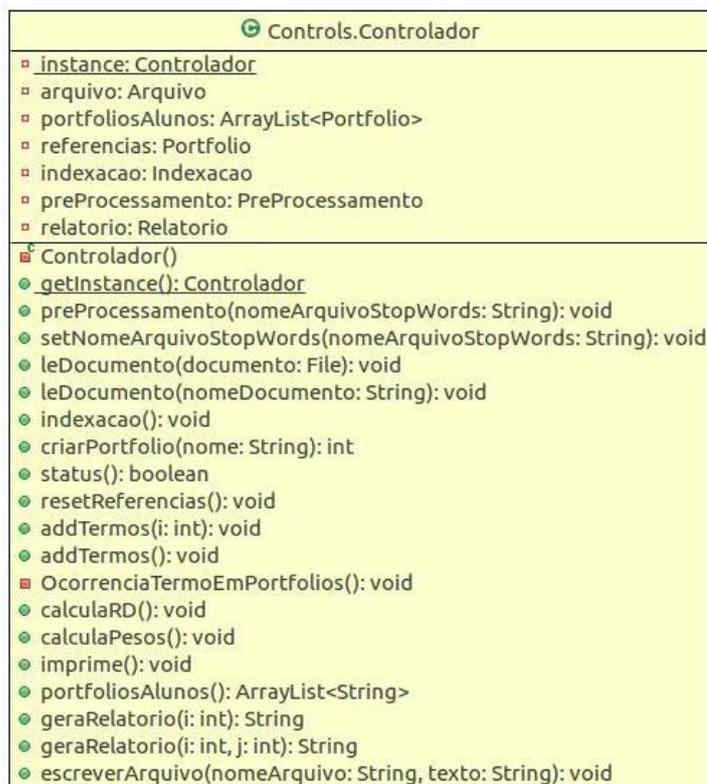


Figura 14: Diagrama de Classe – Classe Controlador. Imagem gerada pelo autor.

Alguns dos métodos são basicamente chamadas aos métodos das classes da camada *Model*. Assim, se uma classe da camada *View* quer acessar uma informação de alguma classe da camada *Model*, ela invoca um método específico na classe “Controlador” que busca a informação acessando um método correspondente na camada *Model*. Desta forma, as camadas *View* e *Model* se comunicam de forma indireta através da classe “Controlador” da camada *Controller*.

O objeto *instance* é um objeto do tipo “Controlador” e somente ele que será acessado durante a execução da ferramenta para lidar com as questões de controle da comunicação, criação de objetos e alteração de atributos da camada *Controller*. Ele fica armazenado na própria classe “Controlador”, garantindo que todas as demais partes do sistema irão lidar apenas com um “Controlador”, o próprio objeto *instance*, ou seja não existiram dois objetos do tipo “Controlador” sendo executados ao mesmo tempo no sistema. Quando alguma classe necessita acessar o controlador, ela pode fazê-lo através do método *getInstance()*. Este método retorna uma espécie de canal de acesso ao objeto *instance*, que ao ser invocado pela primeira vez cria um objeto do tipo “Controlador”. O objeto gerado pode ser acesso através de uma espécie de cópia do “Controlador”. Porém este método garante que as cópias geradas estarão

sincronizadas e cada alteração nas informações de alguma das cópias do objeto *instance* as informações serão replicadas automaticamente para as demais.

5. Testes e Resultados

Esta seção apresenta uma análise avaliativa da ferramenta proposta. O foco principal é determinar sua aplicabilidade no apoio a avaliação de portfólios reais produzidos por alunos durante as disciplinas. Portfólios de alunos e materiais de referência foram utilizados como entrada para gerar os relatórios de avaliação nos moldes do que foi descrito na seção 3.4.2, objetivando estimar a eficácia da ferramenta.

Nas próximas seções são apresentadas a descrição do experimento, os resultados obtidos e uma discussão destes resultados.

5.1. Descrição do Experimento

Os testes foram realizados utilizando material de duas disciplinas, “Banco de Dados I” e “Inteligência Artificial”, ambas pertencentes à grade do curso de Sistemas de Informação da Universidade Federal de Ouro Preto – Minas Gerais. O material consiste de artigos sobre assuntos específicos dentro das duas disciplinas e foram gerados por grupos de alunos que receberam notas do professor entre 0 e 10.

Na disciplina de Banco de Dados I o material produzido era oriundo do tema: “Banco de Dados Móveis”. Na disciplina de “Inteligência Artificial os temas abordados foram: “Lógica Fuzzy” e “Robótica”. Todo o material foi disponibilizado no formato “.pdf” e foi transformado para o formato de texto “.txt” por questões de desempenho da ferramenta e devido a simplicidade deste formato, como foi discutido na seção 3.4.1.

A conversão de formatos de texto foi executada através do software “pdftotext” que é nativo em algumas distribuições do Linux. Esse software cria um novo documento no formato “txt” a partir do documento passado como parâmetro e possui configurações como: intervalos de páginas que devem ser convertidas, a codificação

das palavras no arquivo de saída, etc., além de realizar a conversão dos formatos de forma rápida. Como o formato “txt” aceita apenas texto puro, a ferramenta ignora qualquer imagem na conversão, ou seja, já realiza um primeiro tratamento no texto.

Após a conversão, foram removidos dos textos, de forma manual, os cabeçalhos, rodapés, referências bibliográficas, etc., com o intuito de excluir da análise informação indesejada que poderia alterar os resultados. Ao fim deste processo, os documentos foram movidos para o repositório de portfólios e agrupados em pastas de acordo com seu respectivo portfólio.

Foram considerados para a disciplina de “Inteligência Artificial”, 10 documentos produzidos por 5 grupos distintos de alunos. Cada grupo gerou dois artigos, um de cada tema: “Lógica Fuzzy” e “Robótica”. Como os temas são específicos, foi necessário gerar dois portfólios para cada grupo, cada um contendo dois arquivos de referência. A representação do portfólio de referência do tema “Lógica Fuzzy” foi gerada utilizando o trabalho de Aguado e Cantanhede (2010) e trechos do livro de Russell e Norving (2013). De maneira análoga, o portfólio de referência do tema “Robótica” foi gerado através do trabalho de Gudwin (2005) e trechos do livro de Russell e Norving (2013).

Para a disciplina de Banco de Dados, o teste foi realizado em três documentos gerados por três grupos distintos de alunos. O tema destes três artigos foi “Banco de Dados Móveis”. Foram gerados três portfólios, um pertencendo a cada grupo. O portfólio de referência foi gerado através do livro Sistemas de Banco de Dados (ELMASRI & NAVATHE 2005).

Na próxima seção são apresentados e discutidos os resultados obtidos com a utilização da ferramenta usando o material citado anteriormente.

5.2. Resultados

Esta seção apresenta os resultados obtidos com os portfólios dos alunos e os portfólios de referência que foram gerados para os temas “Lógica Fuzzy” e “Robótica” da disciplina “Inteligência Artificial” e o tema “Bancos de Dados Móveis” da disciplina de “Banco de Dados”.

Os portfólios de cada aluno foram comparados por intermédio de suas ontologias fuzzy utilizando-se do algoritmo da seção 3.3.1. Essa comparação gerou o “grau de compatibilidade” entre os portfólios analisados, que é uma métrica importante no auxílio à avaliação dos portfólios. Ele é apresentado no “relatório de avaliação” que dispõe também de informações sobre os conceitos mais significativos dentro dos portfólios comparados e os relacionamentos mais importantes dentro de cada portfólio. Na figura a seguir temos o exemplo de um relatório gerado através da ferramenta.

Relatório de Avaliação de Portfólios Data e horário da geração do Relatório: 26/03/2017 04:48:19

Portfólio: Referencias			Ontologia Fuzzy Relacionamentos entre termos:		
Conceitos	Ocorrência total no Portfólio	Peso no portfólio	Termo	Termo	Grau
robos	122	169,128	robos	robo	0,943
robo	115	159,424	robos	robotica	0,795
robotica	97	134,471	robo	robotica	0,783
figura	54	85,153	robos	sensores	0,500
sensores	61	84,564	robo	sensores	0,470

Portfólio: Grupo B			Ontologia Fuzzy Relacionamentos entre termos:		
Conceitos	Ocorrência total no Portfólio	Peso no portfólio	Termo	Termo	Grau
robos	61	96,192	robos	robo	0,639
robo	39	61,500	robos	robotica	0,475
robotica	29	45,731	robos	sistemas	0,246
time	8	31,135	robos	inteligencia	0,230
sistemas	15	23,654	robos	humanos	0,230

Valor do CD com Jaccard 1: 0.542013188725018
 Valor do CD com Jaccard 2: 0.5420139216986384
 Valor do CD com Jaccard 3: 0.5420140719371535
 Valor do CD com Jaccard 4: 0.5420129654351187

Figura 15: Relatório de Avaliação de Portfólio. Imagem gerada pelo autor.

A figura 16 apresenta uma lista com as palavras mais significativas para cada um dos portfólios analisados, uma lista com os relacionamentos entre conceitos mais significativos e mostra a variação do valor do “grau de compatibilidade” que foi calculado com quatro variações do coeficiente de similaridade de Jaccard. De posse destas informações o professor pode realizar análises diferentes com o intuito de avaliar o portfólio do aluno.

Através deste relatório foram executados os testes nos portfólios dos alunos de “Banco de Dados” e “Inteligência Artificial”. Os artigos produzidos pelos alunos receberam notas entre 0 e 10. Para proporcionar uma melhor visualização dos dados nos gráficos, os grupos foram ordenados de forma decrescente à nota obtida na

disciplina. A tabela I mostra as notas fornecidas pelo professor para cada um dos grupos de alunos em seus artigos de Inteligência Artificial, com o tema “Lógica Fuzzy”.

Alunos	Grupo A	Grupo B	Grupo C	Grupo D	Grupo E
Notas	10	8.5	8.5	7.5	7

Tabela I: Notas dos Alunos de Inteligência Artificial no tema Lógica Fuzzy. Tabela Gerada pelo autor.

Estas notas foram contrastadas com o valor do “Grau de Compatibilidade” que foi calculado para cada portfólio dos grupos de alunos, seguindo as métricas estabelecidas na seção 3.3.1. O resultado desta comparação pode ser visto na figura a seguir.

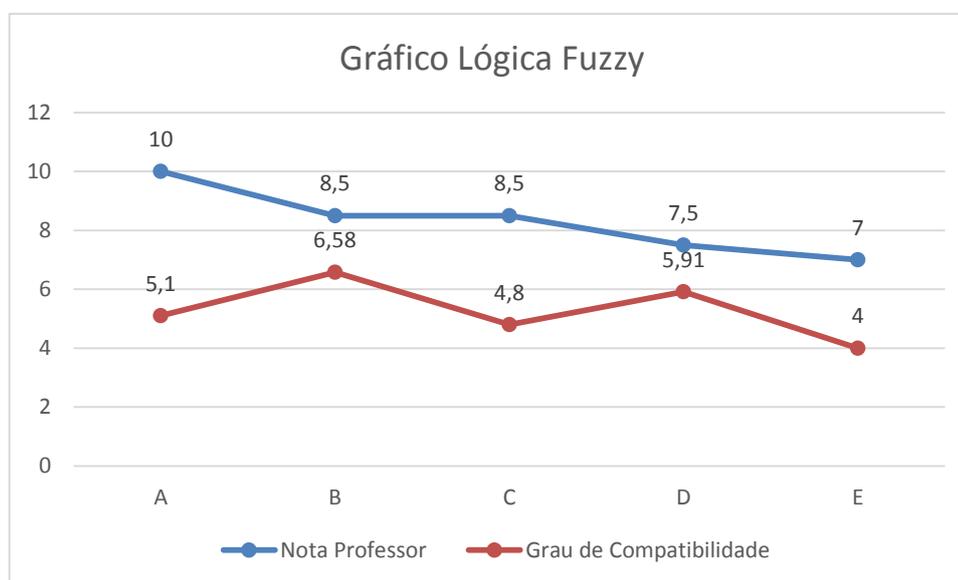


Figura 16: Gráfico de Linhas Comparativo do tema Lógica Fuzzy. Imagem gerada pelo autor.

Na figura 16 é apresentada uma descrição gráfica entre os valores das notas dadas pelo professor e o “grau de compatibilidade” calculado pela ferramenta e disponibilizado no relatório de avaliação. As notas dadas pelo professor aos grupos B e D possuem uma correlação razoavelmente próxima. A discussão detalhada destes valores será apresentada na próxima seção. A seguir temos a tabela II que apresenta as notas obtidas pelos grupos de alunos no tema “Robótica”.

Alunos	Grupo A	Grupo B	Grupo C	Grupo D	Grupo E
Notas	9	9	8.5	8.5	8.5

Tabela II: Notas dos Alunos de Inteligência Artificial no tema Robótica. Tabela gerada pelo autor.

De maneira análoga à tabela I apresentada anteriormente, os valores das notas dos alunos apresentadas na tabela II foram comparados com os valores calculados do “grau de compatibilidade”. Esses valores podem ser vistos na figura a seguir.

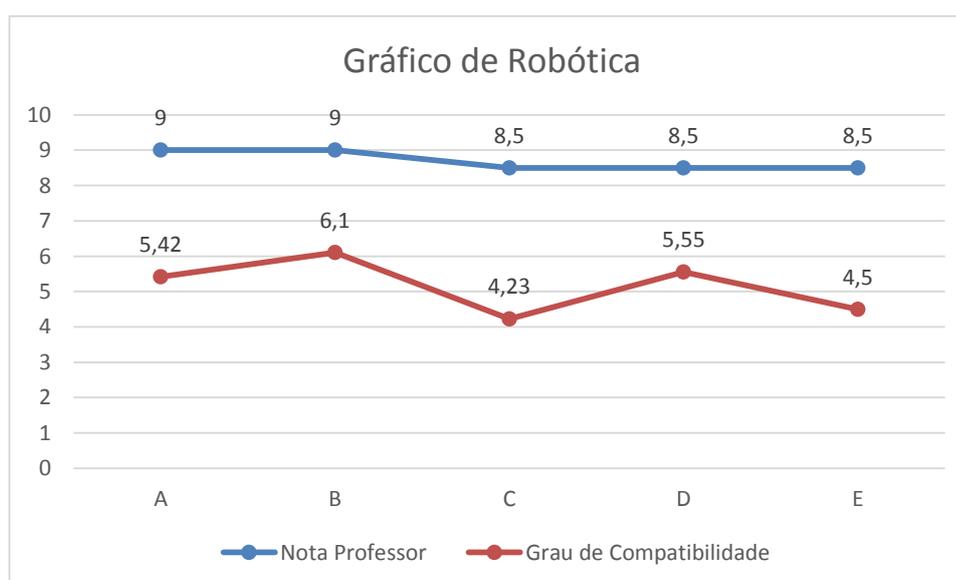


Figura 17: Gráfico de Linhas Comparativo do tema Robótica. Imagem gerada pelo autor.

Observando a figura 17, é possível notar que, apesar de não haver uma correlação direta entre as notas obtidas pelos alunos e o “grau de compatibilidade” fornecida pela ferramenta, existe uma certa relação no que diz respeito à diferença de pontuação entre os diversos grupos. Por exemplo, a diferença entre as notas obtidas pelo Grupo B e pelo Grupo D é praticamente a mesma em ambas as avaliações.

Por fim é apresentado a tabela III com as notas obtidas pelos grupos de alunos na disciplina de “Banco de Dados” para o tema “Bancos de Dados Móveis.

Alunos	Grupo A	Grupo B	Grupo C
Notas	9	8.5	7.5

Tabela III: Notas dos Alunos de Banco de Dados no tema Bancos de Dados Móveis. Tabela gerada pelo autor.

Semelhante ao que foi feito com os dados dos alunos da disciplina “Inteligência Artificial”, esses valores foram contrastados com os valores de “grau de compatibilidade” apresentados no “Relatório de Avaliação” do portfólio de cada grupo e são apresentados graficamente pela figura a seguir.

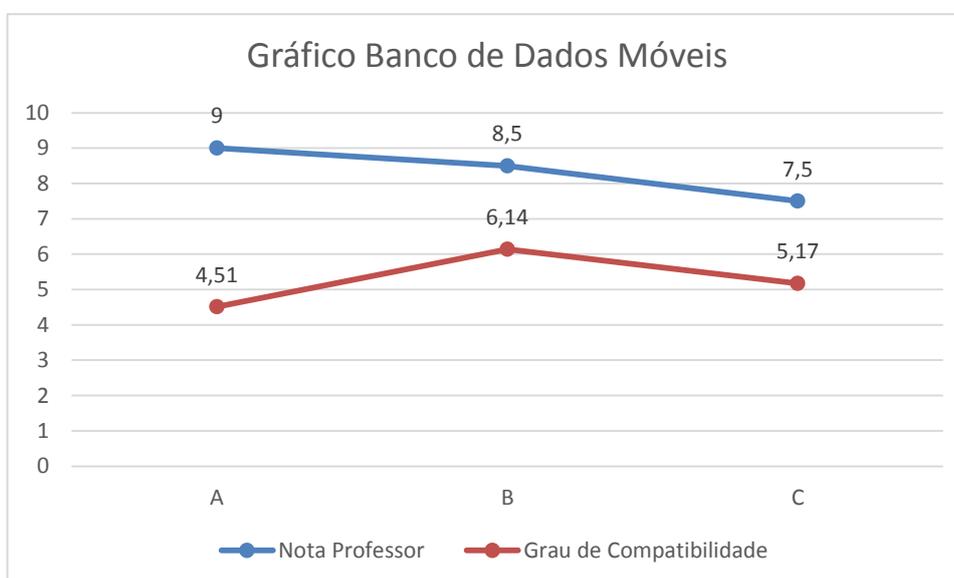


Figura 18: Notas dos Alunos de Banco de Dados no tema Banco de Dados Móveis. Imagem gerada pelo autor.

Para os grupos B e C, observamos uma certa simetria na curva dos valores, onde a diferença das notas é de 1,0 ponto. Outro ponto que chama a atenção é disparidade entre os valores do grupo A. A nota recebida por este grupo é duas vezes maior que o seu respectivo “grau de compatibilidade”.

O que pode ser verificado com os testes é que a construção da base de dados deve ser muito bem executada. Devem ser adicionadas referências com relação próxima aos portfólios dos alunos para que os resultados sejam satisfatórios. De maneira análoga, podem ser executados testes com bases mais genéricas exatamente com o intuito de verificar se os portfólios dos alunos possuem boa construção não fugindo muito ao tema proposto.

Durante a construção dos portfólios de referência, testes foram executados na busca de uma base que representasse, de maneira significativa, os temas analisados. Os resultados apresentados demonstram que a ferramenta consegue representar os portfólios e gerar informações que auxiliam no processo de avaliação dos portfólios.

Fazendo uma análise mais profunda dos resultados, foi possível verificar que a diferença entre os valores do “grau de compatibilidade” e as notas recebidas pelos alunos pode ser explicada por três fatores principais. Em primeiro lugar, os artigos produzidos pelos alunos geralmente estendem o tema central objeto de investigação que foi proposto pelo professor. Isso pode acarretar na incorporação de assuntos mais genéricos dentro dos documentos produzidos, reduzindo a compatibilidade com o portfólio de referência. Em segundo lugar, a nota atribuída ao aluno pelo professor é guiada por métricas subjetivas que determinam o nível de conformidade com os objetivos do trabalho proposto. Estas métricas podem levar em conta outros fatores que afetam as notas, tais como a pontualidade na entrega dos trabalhos, clareza na apresentação das ideias nos documentos, etc. Em terceiro lugar, a própria construção do portfólio de referências pode alterar significativamente o “grau de compatibilidade” entre os portfólios. Se ele for genérico demais, pode superestimar este valor; por outro lado, se ele for específico demais pode gerar um “grau de compatibilidade” abaixo do

esperado. Devido a estes três fatores o “grau de compatibilidade” e a nota do aluno podem ser diferentes.

Duas ressalvas devem ser feitas com a relação à estrutura proposta. Em primeiro lugar, o “grau de compatibilidade” é apenas uma estimativa da proximidade entre o portfólio de referência e o portfólio do aluno, não devendo ser utilizado como nota final do aluno. Em segundo lugar, o relatório de avaliação fornece também, além do “grau de compatibilidade”, informações de conceitos mais significativos dentro do portfólio, com seus respectivos pesos de representatividade, e os relacionamentos entre conceitos mais significativos na construção do portfólio. Estas informações, juntamente com o “grau de compatibilidade”, servem como guia no processo de avaliação dos portfólios.

Na próxima seção são apresentados os trabalhos futuros e a conclusão do trabalho.

6. Conclusão e Trabalhos Futuros

A utilização de portfólios como ferramenta pedagógica vem ganhando espaço principalmente no meio acadêmico. Contudo os meios de avaliação dos portfólios não são muito explorados e este processo acaba sendo realizado de forma manual. Neste contexto, o presente trabalho propôs um método de apoio na avaliação de portfólios, utilizando-se de ontologias fuzzy para representar semanticamente o conteúdo do portfólio.

A estrutura proposta se mostrou flexível e apresentou um poder representacional elevado, o que permite diferentes análises. O relatório de portfólio gerado pela ferramenta utiliza de uma estratégia de similaridade aplicada ao portfólio de referência e ao portfólio de um aluno. As informações geradas por este relatório ajudam na tarefa de avaliação dos portfólios. Esta solução é uma abordagem que visa minimizar o esforço demandado pela avaliação de portfólios e não substituir a avaliação humana.

O experimento realizado utilizou um repositório de portfólios e uma coleção de documentos relacionados com as disciplinas mencionadas na seção anterior. Os resultados obtidos mostram que a ferramenta proposta pode ser útil na tarefa de

avaliar o conteúdo dos portfólios. Por outro lado, a ferramenta necessita que o portfólio de referência seja configurado adequadamente, de modo que este não seja nem específico demais, nem muito genérico.

Uma proposta de melhoria na ferramenta está na utilização de técnicas de “stemming” de termos na fase de pré-processamento. Esta fase faz parte do “Processamento de Linguagem Natural” que foi apresentado na seção 2.4 deste trabalho. O “stemming”, a exemplo da remoção de “stopwords”, gera uma redução de vocabulário, tornando os “termos” mais específicos. Além disso, experimentos mais detalhados com uma maior quantidade de dados também se faz necessária.

As pesquisas realizadas neste trabalho foram dificultadas e limitadas pela criação de uma representação do portfólio de referência de cada assunto abordado nos testes, onde cada representação de portfólio de referência foi criada a partir de artigos referentes ao assunto abordado pelos trabalhos realizados pelos alunos. Esse fato demandou um maior esforço para gerar estes portfólios de referência e impactou no “Grau de compatibilidade” calculado, não gerando um resultado muito próximo das notas obtidas pelos alunos.

A criação de uma ferramenta similar em uma linguagem web pode dar maior adesão de uso deste mecanismo de auxílio à avaliação de portfólios, devido a sua facilidade de propagação e acesso dos usuários ao redor do mundo. Outra consideração importante seria a incorporação de técnicas de inteligência artificial para realizar análises sintáticas do conteúdo dos portfólios, pois a ferramenta proposta neste trabalho realiza apenas a análise do conteúdo semântico do texto, não avaliando a construção das frases e parágrafos dos documentos, somente o conjunto de palavras em um contexto separado. Estas alterações podem dar a ferramenta maior poder representacional, além de gerar análises mais precisas, o que pode melhorar os resultados obtidos facilitando ainda mais a tarefa de avaliação dos portfólios.

Referências

- AGUADO, A. G.; CANTANHEDE, M. A. Lógica Fuzzy. **Faculdade de Tecnologia da UniCAMP**, p. 12, 2010. Disponível em: <http://www.ft.unicamp.br/liag/wp/monografias/monografias/2010_IA_FT_UNICAMP_logicaFuzzi.pdf>.
- ALMEIDA, M. B.; BAX, M. P. Uma visão geral sobre ontologias: pesquisa sobre definições, tipos, aplicações, métodos de avaliação e de construção. **Ciência da Informação**, v. 32, n. 3, p. 7–20, 2003.
- ALMEIDA, M. L. de. Elementos Finitos Paramétricos Implementados em Java. 2005.
- ALVARENGA, G. M.; ARAUJO, Z. R. Portfólio: conceitos básicos e indicações para utilização. **Estudos em Avaliação Educacional**, v. 17, n. 3, p. 137–148, 2006a. Disponível em: <<http://www.fcc.org.br/pesquisa/publicacoes/eae/arquivos/1281/1281.pdf>>.
- ALVARENGA, G. M.; ARAUJO, Z. R. Portfólio: aproximando o saber e a experiência. **Estudos em Avaliação Educacional**, v. 17, n. 34, p. 187–206, 2006b.
- AYRES, R. M. J. Mineração de Regras de Associação Generalizadas Utilizando Ontologias Fuzzy e Similaridades Baseadas em Contexto. 2012.
- BARROS, R. S. M. Medidas de diversidade biológica. **Programa de Pós-Graduação em Ecologia Aplicada ao Manejo e Conservação de Recursos Naturais – PGEOL. Universidade Federal de Juiz de Fora – UFJF.**, p. 13, 2007.
- BEPPLER, F. D. Um modelo para recuperação e busca de informação baseado em ontologia e no círculo hermenêutico. 2008. Disponível em: <<https://repositorio.ufsc.br/handle/123456789/90972>>.
- CHEN, C.; CHEN, Y. Learning performance assessment approach using learning portfolio for e-learning systems. **Fifth IEEE International Conference on Advanced Learning Technologies (ICALT'05)**, p. 557–561, 2005. Disponível em: <<http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=1508756>>.
- DUARTE, K. C.; FALBO, R. de A. Uma ontologia de qualidade de software. **Workshop de Qualidade de Software**, 2000. Disponível em: <<http://inf.ufes.br/~falbo/download/pub/Wqs2000.pdf>>.
- ELMASRI, R.; NAVATHE, S. B. **Sistema de banco de dados**. [s.l.: s.n.]
- FOURALI, C. Using Fuzzy Logic in Educational Measurement: The Case of Portfolio Assessment. **Evaluation & Research in Education**, v. 11, n. 3, p. 129–148, 1997.
- GOMES, A. P.; ARCURI, M. B.; CRISTEL, E. C.; RIBEIRO, R. M.; SOUZA, L. M. B. da M.; SIQUEIRA-BATISTA, R. Avaliação no Ensino Médico: o papel do portfólio nos currículos baseados em metodologias ativas. **Revista Brasileira de Educação Médica**, v. 34, n. 3, p. 390–396, 2010. Disponível em: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0100-55022010000300008&lng=pt&nrm=iso&tlng=pt>.
- GOMIDE, F. A. C.; GUDWIN, R. R. Modelagem, Controle, Sistemas E Lógica Fuzzy. v. 4, 1994.
- GOMIDE, F. A. C.; GUDWIN, R. R.; TANSCHKEIT, R. Conceitos fundamentais da teoria de conjuntos fuzzy, lógica fuzzy e aplicações. **Proc. 6 th IFSA Congress-Tutorials**, p. 1–38, 1995.
- GONÇALVES, M. A. B. GERAÇÃO DE MALHAS BIDIMENSIONAIS DE ELEMENTOS FINITOS BASEADA EM MAPEAMENTOS TRANSFINITOS. 2004.
- GONZALEZ, M.; LIMA, V. L. S. de. Recuperação de Informação e Processamento da Linguagem Natural. **XXIII Congresso da Sociedade Brasileira de Computação. Anais da III Jornada de Mini-Cursos de Inteligência Artificial. Campinas:[sn]**, v. 3, p. 347–395, 2003. Disponível em: <http://www.erfelipe.com.br/artigos/RI_Processamento_de_linguagem_natural.pdf>.
- GUDWIN, R. R. Novas Fronteiras na Inteligência Artificial e na Robótica Resumo: Introdução Robótica Industrial e Algoritmos de Planejamento Robôs Móveis. p. 1–18, 2005.
- INDRUSIAK, L. S. Linguagem java. **Grupo JavaRS JUG Rio Grande do Sul**, 1996. Disponível em: <<http://www.cin.ufpe.br/~arfs/introjjava.pdf>>.
- INE. **Portal do Consórcio World Wide Web no Brasil**. Disponível em: <<http://www.w3c.br/Home/WebHome>>. Acesso em: 3 fev. 2016a.

- INE. **A linguagem JAVA**. Disponível em: <https://www.java.com/pt_BR/about/>. Acesso em: 6 fev. 2016b.
- LARANJEIRO, M. A.; COSTA, F. A. E-Portfolio - in - Education Practices and Reflections. 2008.
- MEYER, A. D. S. Comparação de Coeficientes de Similaridade usados em Análises de Agrupamento com Dados de Marcadores Moleculares Dominantes. **Usp**, p. 106, 2002.
- MINETTO, E. L. Frameworks para Desenvolvimento em PHP. p. 192, 2007.
- MORAES, O. B. de. Método de análise de dados para avaliação de áreas urbanas recuperadas – uma abordagem utilizando a lógica fuzzy. 2008.
- PEREIRA, R.; RICARTE, I. L. M.; GOMIDE, F. Ontologia relacional fuzzy em sistemas de recuperação de informação. **Congresso da Sociedade Brasileira de Computação**, n. FEBRUARY 2005, p. 672–681, 2005.
- PILLAR, V. de P. Variações espaciais e temporais na vegetação; métodos analíticos. **Departamento de Botânica UFRGS**, n. Orlóci 1991, p. 14, 1996. Disponível em: <http://ecoqua.ecologia.ufrgs.br/arquivos/Reprints&Manuscripts/Manuscripts&Misc/5_MetodosAnaliticos_96Out14.pdf>.
- RUSSELL, S.; NORVING, P. **Inteligência Artificial - Tradução da terceira edição**. [s.l.: s.n.]
- SANDRI, S.; CORREA, C. Lógica Nebulosa. **The Korean Academic Year Fall Conference 2006**, p. 1–14, 2006. Disponível em: <[http://www.papersearch.net/google_link/fulltext.asp?file_name=04710746.pdf%5Cfile:///Users/superman/Library/Application Support/Papers2/Articles/2006/???/2006%20.pdf%5Cpapers2://publication/uuid/C5C4C08C-239B-4A3A-AA8E-29DA0BBD9226](http://www.papersearch.net/google_link/fulltext.asp?file_name=04710746.pdf%5Cfile:///Users/superman/Library/Application%20Support/Papers2/Articles/2006/???/2006%20.pdf%5Cpapers2://publication/uuid/C5C4C08C-239B-4A3A-AA8E-29DA0BBD9226)>.
- SATLER, M. F. Perfiles de Usuario Basados en Ontologías Borrosas para la Mejora de las Tareas de Filtrado de Información. 2011.
- SATLER, M. F.; CASTRO, C. V.; ROMERO, F. P.; OLIVAS, J. A.; BRAGA, J. L. Enhancing portfolio assessment: An application of fuzzy ontologies. **Intelligent Systems Design and Applications (ISDA), 2011 11th International Conference on**, p. 990–995, 2011.
- SILVA, L. F. S.; CHAPETTA, W. A. Processamento de Linguagem Natural Aplicada à Inspeção de Documentos de Especificação de Requisitos de Software. **Antimicrobial Agents and Chemotherapy**, v. 53, n. 12, p. 5026–34, 2003. Disponível em: <<http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2786324&tool=pmcentrez&rendertype=abstract>>.
- SOUZA, D. M. B. de. Abordagem Baseada em Lógica Fuzzy para Alocação de Indicadores de Faltas em Sistemas de Distribuição de Energia Elétrica. 2009.
- VIANA, M. C. Uma Proposta de Reúso de Interface Gráfica com o Usuário Baseada no Padrão Arquitetural MVC. n. Dc, 2008.
- WIVES, L. K. Um estudo sobre agrupamento de documentos textuais em processamento de informações não estruturadas usando técnicas de “clustering”. 1999. Disponível em: <<http://www.lume.ufrgs.br/handle/10183/25933>>.
- ZADEH, L. A. **Fuzzy Sets-Information and Control-1965**, 1965. .