
Abordagem de Recomendação baseada
em Conteúdo utilizando Ontologia
Fuzzy de Domínio e Ontologia Crisp de
Preferência do Usuário

Arturo Elias Urquizo Baldárrago

PROGRAMA DE PÓS-GRADUAÇÃO DA UFSCar

Data de Depósito: 15/09/2012

Assinatura: _____

Abordagem de Recomendação baseada em Conteúdo utilizando Ontologia *Fuzzy* de Domínio e Ontologia Crisp de Preferência do Usuário

Arturo Elias Urquizo Baldárrago

Orientador: *Prof. Dr. Antonio Francisco Prado*

Dissertação apresentada ao Programa de Pós-Graduação em Ciência da Computação da Universidade Federal de São Carlos, como parte dos requisitos para a obtenção do título de Mestre em Ciência da Computação, área de concentração: Engenharia de Software e Banco de Dados.

UFSCar - São Carlos
Setembro/2012

**Ficha catalográfica elaborada pelo DePT da
Biblioteca Comunitária da UFSCar**

U79ar

Urquizo Baldárrago, Arturo Elias.

Abordagem de recomendação baseada em conteúdo utilizando ontologia fuzzy de domínio e ontologia crisp de preferência do usuário / Arturo Elias Urquizo Baldárrago. -- São Carlos : UFSCar, 2012.

152 f.

Dissertação (Mestrado) -- Universidade Federal de São Carlos, 2012.

1. Banco de dados. 2. Ontologia difusa. 3. Ontologia de usuário. 4. Sistemas de recomendação. I. Título.

CDD: 005.74 (20ª)

Universidade Federal de São Carlos
Centro de Ciências Exatas e de Tecnologia
Programa de Pós-Graduação em Ciência da Computação

**“Abordagem de Recomendação baseada em
Conteúdo utilizando Ontologia Fuzzy de Domínio
e Ontologia Crisp de Preferência do Usuário”**

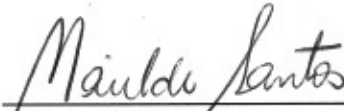
Arturo Elías Urquizo Baldárrago

Dissertação de Mestrado apresentada ao
Programa de Pós-Graduação em Ciência da
Computação da Universidade Federal de São
Carlos, como parte dos requisitos para a
obtenção do título de Mestre em Ciência da
Computação


Membros da Banca:



Prof. Dr. Antonio Francisco do Prado
(Orientador - DC/UFSCar)



Prof. Dra. Marilde Terezinha Prado Santos
(DC/UFSCar)



Prof. Dr. Carlos Roberto Valêncio
(UNESP)

São Carlos
Julho/2012

Agradecimentos

Agradeço a Deus, que permitiu que eu chegasse neste ponto, também por ter colocado anjos na minha vida que me ajudaram a atravessar esse difícil caminho.

Aos meus pais, Arturo e Hilda, e à minha irmã Karito, que a cada final de semana demonstraram todo o amor e carinho que sentem por mim. Também agradeço pelo apoio que me deram e mesmo distantes, sempre estiveram presentes em pensamentos e orações. Amo vocês!!! Muito obrigado!!!

Aos meus avós Amelia, Julieta, Elias e Arturo, que, apesar da distância, sempre estiveram presentes em pensamentos e orações.

À minha namorada, Nathalie, que mais que uma companheira é o complemento da minha vida. Agradeço-a infinitamente por todo apoio, e estou seguro de que sem ela não teria conseguido chegar aqui. Obrigado por todos os sacrifícios que fizestes para dedicar-me um pouco de seu tempo. Também agradeço por ter compartilhado comigo tantos momentos felizes que me ajudaram a seguir adiante com ânimo. Simplesmente te amooo...ooo (infinitos o's). Muito obrigado por não me deixar desanimar!!!

Ao meu orientador, professor Prado, e à minha coorientadora, professora Marilde, que me deram essa oportunidade e confiaram em mim. Também agradeço por todo apoio que me ofereceram ao longo desses anos.

Aos meus amigos Merley, Paula, Victor, Dú, DB, Paulo, Mayra, Augusto e Thiago, por terem compartilhado comigo tantos momentos, rindo, reclamando, celebrando e até chorando. Galera, vocês são 10!!!!

Aos meus “filhos”, Motitas, Blackito, Rayito e Leehlah, por todos os momentos de diversão e de alegria que me proporcionaram toda vez que chegava em casa, fazendo com que as frustrações do dia desaparecessem.

Aos colegas e amigos que vieram comigo do Perú, e àqueles que fiz amizade estando aqui. Obrigado pelos momentos de descontração que me proporcionaram.

Aos colegas e amigos do PPG-CC, com os quais compartilhei muitas coisas durante o percurso do mestrado, pelos incentivos mútuos para seguir adiante.

Aos colegas e amigos do LaBDES, pelo companheirismo mostrado em cada momento, assim como por todos os ensinamentos de português durante “a hora do café”. Obrigado, galera!

Aos colegas e amigos do LABIC-USP, que me ajudaram em várias etapas do

ii

mestrado e que compartilharam comigo momentos agradáveis. Obrigado!

À CAPES e ao PPG-CC, pelo apoio financeiro

Enfim, a todos que, de alguma maneira, tiveram uma passagem na minha vida durante esse período e contribuíram para a realização deste trabalho. Que Deus abençoe a todos.

Muito Obrigado!

Resumo

Esta dissertação apresenta uma abordagem para o desenvolvimento de aplicações de recomendação baseadas em conteúdo utilizando ontologia específica de domínio e ontologia de preferência de usuário. Tal abordagem está dividida em duas etapas: a Engenharia de Ontologia e a Engenharia do Sistema de Recomendação. Na Engenharia de Ontologia são construídas: uma ontologia de domínio com relacionamentos difusos; e uma ontologia crisp de usuário definida como uma instância da ontologia de domínio, porém modelada de forma que permita refletir as preferências de cada usuário para o domínio instanciado. A utilização das ontologias produzidas na Engenharia de Ontologia proporciona um ganho de precisão nos resultados obtidos por aplicações desenvolvidas conforme a abordagem proposta. Para fins de avaliação, a abordagem proposta foi instanciada no domínio de comércio de dispositivos móveis. Seguindo a metodologia experimental, foi conduzida uma experimentação com o objetivo de avaliar o impacto da abordagem na precisão dos resultados fornecidos pelo Sistema de Recomendação. Os resultados evidenciaram que o uso da abordagem proposta colaborou para o incremento da precisão dos resultados. As contribuições deste trabalho incluem: a abordagem para o desenvolvimento de aplicações de recomendação baseadas em conteúdo utilizando ontologia *fuzzy* específica de domínio e ontologia de preferência de usuário; a definição da metodologia de construção de ontologias *fuzzy* chamada UPFON; a instanciação de uma ontologia *fuzzy* no domínio dos dispositivos móveis e a estratégia para capturar as preferências do usuário e propagá-las em uma ontologia crisp de usuário.

Palavras-chave: Ontologias Fuzzy, Ontologias de Usuário, Sistemas de Recomendação Baseados em Conteúdo

Abstract

This paper presents an approach for developing content-based recommendation applications with focus on the use of a specific domain fuzzy ontology along with a user preference ontology. The approach falls into two stages: Ontology Engineering and Recommendation System Engineering. In the Ontology Engineering, a domain ontology with fuzzy relationships and a user ontology are built. The user ontology is set as an instance of the domain ontology, but it is modeled in a way that allows to store each user's preferences. The usage of the ontologies produced in Ontology Engineering provides a gain in precision for the results obtained by applications in the Recommendation System Engineering stage. For evaluation purposes, we instantiated the proposed approach in the development of a Recommender System for the field of electronic commerce, focusing on the mobile devices commerce domain. Following the experimental methodology, An evaluation was conducted in order to assess the approach's impact on the accuracy of results provided by the developed Recommender System. The results showed that the use of our approach contributed to increase the accuracy of the results, in terms of prediction, classification and ranking. The contributions of this work include: the approach for developing content-based recommendation applications by using a specific domain fuzzy ontology along with a user preference ontology; the definition of the UPFON methodology, which integrates the approach, to construct fuzzy ontologies; an instantiation of a fuzzy ontology for the mobile devices domain and a strategy to capture; and propagate the user preferences by means of ontologies.

Keywords: Fuzzy Ontologies, User Ontologies, Content Based Recommender Systems.

Sumário

Resumo	iii
Abstract	v
Sumário	vii
Lista de Figuras	ix
Lista de Tabelas	xi
1 Introdução	1
1.1 Contextualização	1
1.2 Objetivos e Hipótese	3
1.3 Principais Resultados	4
1.4 Organização do Trabalho	5
2 Fundamentação Teórica	7
2.1 Mineração de Textos	7
2.1.1 Identificação do Problema	8
2.1.2 Pré-processamento	9
2.1.3 Extração de Padrões usando Agrupamento de Dados	11
2.1.4 Pós-processamento e Utilização do conhecimento	15
2.2 Lógica Fuzzy	15
2.2.1 Noções de lógica <i>Fuzzy</i>	16
2.2.2 Funções de Pertinência	19
2.3 Ontologias	20
2.3.1 Tipos de Ontologias	22
2.3.2 Metodologias para a construção de Ontologias	23
2.3.3 Engenharia da ontologia	29
2.4 Sistemas de Recomendação	30
2.4.1 Processo de Recomendação	32
2.4.2 Métricas de Avaliação	36
2.5 Considerações Finais	38
3 Trabalhos Correlatos	39
3.1 Metodologias para Construção de Ontologias <i>Fuzzy</i>	40
3.1.1 Fuzzificação de uma ontologia tradicional CRISP (Parry, 2004)	40

3.1.2	FOGA - <i>Fuzzy Ontology Generation framework</i> (Quan et al., 2004)	41
3.1.3	Conhecimento <i>fuzzy</i> para integrar ontologias de domínio (Todorov et al., 2011)	42
3.1.4	Roles de usuários na fuzzificação de ontologias (Wallace et al., 2011)	42
3.2	Sistemas de Recomendação baseados em Ontologia de Usuário e Propagação do Interesse	43
3.2.1	Aprendendo e inferindo uma ontologia de usuário para serviços semânticos web personalizados (Jiang e Tan, 2006)	43
3.2.2	Personalização dos mecanismos de busca com perfis ontológicos de usuários (Sieg et al., 2007)	44
3.2.3	Perfil de usuário ontológico em recomendações personalizadas em <i>e-Commerce</i> (He e Fang, 2008)	45
3.2.4	Busca de informação contextual baseada no perfil ontológico de usuário (Mohammed et al., 2010)	46
3.2.5	Abordagem de ontologia <i>fuzzy</i> para representar perfis de usuário em ambientes de <i>e-learning</i> (Satler et al., 2010)	46
3.3	Considerações Finais	48
4	ObRA: Uma Abordagem para Recomendação baseada em Ontologias de Domínio e em Preferências de Usuário	49
4.1	Visão Geral da Abordagem ObRA	50
4.1.1	Engenharia de Ontologia	51
4.1.2	Engenharia do Sistema de Recomendação	62
4.2	Estudo de Caso: Um Sistema de Recomendação Conforme a Abordagem ObRA	64
4.2.1	Engenharia de Ontologia: Ontologia <i>Fuzzy</i> do Domínio de Dispositivos Móveis e Ontologia Crisp de Preferências de Usuário	64
4.2.2	Engenharia do Sistema de Recomendação	69
4.3	Avaliação da Estratégia de Propagação de preferência do usuário	74
4.4	Considerações Finais	77
5	Avaliação Experimental da Abordagem ObRA	79
5.1	Experimentação da Abordagem ObRA	79
5.2	Desenvolvimento da Experimentação	80
5.2.1	Definição do Experimento	80
5.2.2	Planejamento do Experimento	81
5.2.3	Execução do Experimento	83
5.3	Análise e Interpretação dos Resultados	90
5.3.1	Teste das Hipóteses	92
5.3.2	Conclusões do Experimento	96
5.3.3	Ameaças à Validade	96
5.4	Considerações Finais	98

6 Conclusão	101
6.1 Principais Resultados e Contribuições	102
6.2 Principais Limitações	103
6.3 Trabalhos Futuros	103
Referências Bibliográficas	115
A Formulação de Avaliação de Similaridade	117
B Características mais representativas de cada dispositivo	125
C Valores de Erro Absoluto (MAE)	131
D Cálculos de Precisão e Cobertura média	133
E Cálculo e Representação gráfica da Precisão, cobertura e <i>F-score</i> para cada abordagem	137
F Tabela Padrão de Distribuição de Probabilidade Estatística T de Student (Gosset, 1947)	147

Lista de Figuras

1.1	Organização da Dissertação	6
2.1	Etapas do processo de Mineração de Textos (Rezende, 2003)	8
2.2	Etapas do pré-processamento de Textos. Adaptado de Moura (2006)	10
2.3	Representação da matriz atributo-valor para N documentos e M atributos	11
2.4	Exemplo de agrupamento de documentos particional (Rossi, 2011)	13
2.5	Exemplo de um dendrograma obtido por um agrupamento de documentos hierárquico (Rossi, 2011)	14
2.6	Domínio das variáveis linguísticas	16
2.7	Domínio da lógica clássica	16
2.8	Classificação Binária entre ser alto ou não-alto (Shaw e Simões, 1998)	17
2.9	Cesto com bolinhas brancas e escuras adaptada de Pinho (1999)	17
2.10	Cesto com bolinhas variadas (Pinho, 1999)	18
2.11	Funções de pertinência para a variável T. Adaptado de Costa et al. (2007)	19
2.12	Comparação entre função característica e função de pertinência para o conjunto de pessoas altas (Moraes e Marley, 2007)	20
2.13	Comparação entre ontologia e base de conhecimento. Adaptado de Guimarães (2002)	21
2.14	Representação da metodologia UPON de De Nicola et al. (2009), ressaltando os fluxos adaptados para permitir a criação ou fuz-zificação de uma ontologia	25
2.15	Processo de Recomendação	32
4.1	Visão Geral em alto nível da Abordagem de Recomendação baseada em Ontologias <i>Fuzzy</i>	50
4.2	Visão Geral da Engenharia de Ontologias	52
4.3	Representação da metodologia UPON de De Nicola et al. (2009), ressaltando os fluxos adaptados para permitir a criação ou fuz-zificação de uma ontologia	54

4.4	Atividades do Fluxo de Requisitos. a) Atividades da Metodologia UPON. b) Atividades da Adaptação UPFON	56
4.5	Atividades do Fluxo de Análise. a) Atividades da Metodologia UPON. b) Atividades da Adaptação UPFON	57
4.6	Atividades do Fluxo de Projeto. a) Atividades da Metodologia UPON. b) Atividades da Adaptação UPFON	58
4.7	Representação Abstrata de um Relacionamento Binário Difuso	62
4.8	Visão Geral da Engenharia do Sistema de Recomendação	63
4.9	Trecho da ontologia no domínio de dispositivos móveis, utilizada para o estudo	65
4.10	Trecho do repositório WURFL, indicando as características correspondentes ao dispositivo Samsung_X830	65
4.11	Representação abstrata do relacionamento de similaridade utilizando a meta-ontologia <i>fuzzy</i> proposta por Yaguinuma et al. (2010); Cerri (2010)	67
4.12	Representação formal da ontologia <i>fuzzy</i> de domínio	68
4.13	Representação gráfica de uma parte da ontologia <i>crisp</i> de preferência de usuário	69
4.14	Representação formal de uma parte da ontologia <i>crisp</i> de preferência de usuário	69
4.15	Parte do Modelo de Classes como especificação de Análise	70
4.16	Representação Modelo Vista Controlador como parte da especificação de Projeto	70
4.17	Visão geral da arquitetura da aplicação	71
4.18	Trecho do código para acesso aos dados da ontologia	72
4.19	Trecho do código para a propagação do interesse	72
5.1	Cálculo do valor de erro absoluto em relação às avaliações do usuário 01 sobre o primeiro conjunto de dispositivos	87
5.2	Cálculo dos valores de precisão e cobertura média em relação ao primeiro conjunto de dispositivos	88
5.3	Cálculo e representação gráfica dos valores de precisão, cobertura e F-measure para cada uma das abordagens, em relação à avaliação do usuário 01 sobre o primeiro conjunto de dispositivos	89
5.4	Cálculo dos valores de Spearman para cada uma das abordagens, em relação à avaliação do usuário 01 sobre o primeiro conjunto de dispositivos	90
E.1	Representação gráfica da precisão, cobertura e <i>F-score</i> do Usuário 1 considerando os <i>cutt-off</i> para a abordagem de recomendação baseada em ontologia e agrupamento	138
E.2	Representação gráfica da precisão, cobertura e <i>F-score</i> do Usuário 2 considerando os <i>cutt-off</i> para a abordagem de recomendação baseada em ontologia e agrupamento	138
E.3	Representação gráfica da precisão, cobertura e <i>F-score</i> do Usuário 3 considerando os <i>cutt-off</i> para a abordagem de recomendação baseada em ontologia e agrupamento	140

E.4	Representação gráfica da precisão, cobertura e <i>F-score</i> do Usuário 4 considerando os <i>cutt-off</i> para a abordagem de recomendação baseada em ontologia e agrupamento	140
E.5	Representação gráfica da precisão, cobertura e <i>F-score</i> do Usuário 5 considerando os <i>cutt-off</i> para a abordagem de recomendação baseada em ontologia e agrupamento	141
E.6	Representação gráfica da precisão, cobertura e <i>F-score</i> do Usuário 6 considerando os <i>cutt-off</i> para a abordagem de recomendação baseada em ontologia e agrupamento	141
E.7	Representação gráfica da precisão, cobertura e <i>F-score</i> do Usuário 7 considerando os <i>cutt-off</i> para a abordagem de recomendação baseada em ontologia e agrupamento	143
E.8	Representação gráfica da precisão, cobertura e <i>F-score</i> do Usuário 8 considerando os <i>cutt-off</i> para a abordagem de recomendação baseada em ontologia e agrupamento	143
E.9	Representação gráfica da precisão, cobertura e <i>F-score</i> do Usuário 9 considerando os <i>cutt-off</i> para a abordagem de recomendação baseada em ontologia e agrupamento	144
E.10	Representação gráfica da precisão, cobertura e <i>F-score</i> do Usuário 10 considerando os <i>cutt-off</i> para a abordagem de recomendação baseada em ontologia e agrupamento	145

Lista de Tabelas

2.1	Tipos de Ontologia (Almeida e Bax, 2003)	24
2.2	Metodologias para construção de ontologias (Almeida e Bax, 2003)	29
4.1	Trecho da Matriz de Similaridade obtida na atividade de determi- nação do grau de relacionamento dos conceitos	66
4.2	Resultados dos interesses do usuário	74
4.3	Dispositivos relevantes ranqueados de acordo com as preferên- cias do usuário e a abordagem de recomendação sem considerar a propagação do interesse	75
4.4	Cálculo das medidas das métricas de precisão da abordagem de recomendação sem propagação para o conjunto de Dispositivos1	75
4.5	Dispositivos relevantes ranqueados de acordo com as preferên- cias do usuário e a abordagem de recomendação considerando a propagação do interesse	76
4.6	Cálculo das medidas das métricas de precisão da abordagem de recomendação com propagação para o conjunto de Dispositivos1	76
5.1	Dados Coletados da Avaliação	84
5.2	Relação dos Dispositivos Envolvidos para o Primeiro Conjunto de Avaliação	85
5.3	Dados Obtidos ao Avaliar os Dispositivos com as duas Aborda- gens de Recomendação	85
5.4	Organização da Primeira etapa do <i>Paired t-test</i> para a precisão na predição (ρ)	94
5.5	Organização da Segunda etapa do <i>Paired t-test</i> para a precisão na classificação (ϑ)	94
5.6	Organização da Terceira etapa do <i>Paired t-test</i> para a precisão no ranqueamento (π)	95
D.1	Cálculo da Precisão, cobertura e <i>F-Score</i> para o Dispositivo1 se- gundo as abordagens baseada em ontologia e baseada em agru- pamento	134

D.2	Cálculo da Precisão, Cobertura e <i>F-Score</i> para o Dispositivo2 segundo as abordagens baseada em ontologia e baseada em agrupamento	134
D.3	Cálculo da Precisão, Cobertura e <i>F-Score</i> para o Dispositivo3 segundo as abordagens baseada em ontologia e baseada em agrupamento	135
D.4	Cálculo da Precisão, Cobertura e <i>F-Score</i> para o Dispositivo4 segundo as abordagens baseada em ontologia e baseada em agrupamento	135
D.5	Cálculo da Precisão, Cobertura e <i>F-Score</i> para o Dispositivo5 segundo as abordagens baseada em ontologia e baseada em agrupamento	136
D.6	Cálculo da Precisão, Cobertura e <i>F-Score</i> para o Dispositivo6 segundo as abordagens baseada em ontologia e baseada em agrupamento	136
E.1	Cálculos de Precisão, cobertura e F-score para o Usuário 1 considerando a abordagem de recomendação com ontologia e com agrupamento	138
E.2	Cálculos de Precisão, cobertura e F-score para o Usuário 2 considerando a abordagem de recomendação com ontologia e com agrupamento	139
E.3	Cálculos de Precisão, cobertura e F-score para o Usuário 3 considerando a abordagem de recomendação com ontologia e com agrupamento	139
E.4	Cálculos de Precisão, cobertura e F-score para o Usuário 4 considerando a abordagem de recomendação com ontologia e com agrupamento	140
E.5	Cálculos de Precisão, cobertura e F-score para o Usuário 5 considerando a abordagem de recomendação com ontologia e com agrupamento	141
E.6	Cálculos de Precisão, cobertura e F-score para o Usuário 6 considerando a abordagem de recomendação com ontologia e com agrupamento	142
E.7	Cálculos de Precisão, cobertura e F-score para o Usuário 7 considerando a abordagem de recomendação com ontologia e com agrupamento	142
E.8	Cálculos de Precisão, cobertura e F-score para o Usuário 8 considerando a abordagem de recomendação com ontologia e com agrupamento	143
E.9	Cálculos de Precisão, cobertura e F-score para o Usuário 9 considerando a abordagem de recomendação com ontologia e com agrupamento	144
E.10	Cálculos de Precisão, cobertura e F-score para o Usuário 10 considerando a abordagem de recomendação com ontologia e com agrupamento	145

Introdução

1.1 Contextualização

A crescente disponibilidade de dados tem facilitado o acesso à informação pertencente a diversas áreas do conhecimento. Diante da oferta abundante de dados, são necessárias técnicas eficazes para recuperar a informação desejada, com o intuito de reduzir a quantidade de respostas irrelevantes (alta precisão) e assegurar que resultados relevantes não sejam desprezados (alta cobertura) (Khan et al., 2004). Um problema inerente nesse contexto é a sobrecarga de informações. Os usuários ficam com a difícil tarefa de selecionar, a partir de um repositório de informação crescente, os itens que satisfazem seus interesses e necessidades imediatas.

Esse problema é contornado por meio de aplicações capazes de reconhecer o interesse do usuário e o auxiliar em sua consulta. Essas aplicações filtram as informações irrelevantes e classificam os itens encontrados por ordem de relevância, além de sugerir itens que provavelmente são de interesse do usuário, mas que não satisfazem a sua consulta. Esses tipos de aplicação são denominados Sistemas de Recomendação (Ricci et al., 2011).

Um Sistema de Recomendação consiste em um conjunto de técnicas e ferramentas de software que permite sugerir itens que serão de interesse para o usuário (Mahmood e Ricci, 2009; Resnick e Varian, 1997; Burke, 2007). O estudo dos Sistemas de Recomendação é relativamente novo em comparação às pesquisas de outras técnicas clássicas de sistemas de informação (por exem-

plo, bancos de dados ou motores de busca). Nos últimos anos, o interesse em sistemas de recomendação aumentou drasticamente devido ao importante papel que esses desempenham em diversos sítios bem conhecidos da web, como, *YouTube*¹, *Amazon*², *Netflix*³, entre outros.

Uma das técnicas de filtragem utilizadas nos sistemas de recomendação é a filtragem baseada em conteúdo, na qual o sistema aprende a recomendar itens que são semelhantes àqueles que o usuário demonstrou preferência no passado. Uma das grandes dificuldades enfrentadas por esse mecanismo de filtragem é a falta de estruturação da informação na Web. Problemas como a ambiguidade da linguagem natural dificulta o cálculo da similaridade entre um dado elemento de informação e um perfil de usuário, o que pode resultar em itens pouco relevantes após a filtragem (Lops et al., 2011).

Isso acontece porque a Web tradicional foi desenvolvida com foco na maneira como a informação é exibida e pouca atenção foi dispensada à maneira como a mesma é estruturada. Em outras palavras, na Web tradicional os dados podem ser perfeitamente lidos e compreendidos por humanos, mas as entidades de software encontram grandes dificuldades para captar a semântica dos recursos disponíveis. Esse problema pode ser superado a partir da introdução de tecnologias da Web Semântica (Berners-Lee et al., 2001), como a utilização de ontologias para a representação estruturada do domínio em sistemas de recomendação (Middleton et al., 2004; Ivanova et al., 2010).

No entanto, como se sabe, as abordagens que utilizam ontologias clássicas para representar o conhecimento obtido não conseguem representar adequadamente o conhecimento impreciso e vago. Como solução para esse problema, pode-se utilizar o conceito de ontologia *fuzzy*, o que permite lidar com esse conhecimento em diversas aplicações do mundo real. Esse tipo de ontologia vem sendo utilizado em diversas áreas do conhecimento, como na mineração de notícias (Lee et al., 2005), no reconhecimento de imagens (Hudelot et al., 2008), entre outros.

Com o objetivo de prover apoio ao desenvolvimento de aplicações de recomendação que utilizem uma filtragem baseada em conteúdo, propõe-se neste trabalho, uma abordagem de recomendação que usa uma ontologia *fuzzy* de domínio e uma ontologia de preferência do usuário, a qual permite a propagação do interesse do mesmo considerando a similaridade entre conceitos estabelecidos na ontologia *fuzzy*.

¹<http://www.youtube.com>

²<http://www.amazon.com>

³<http://www.netflix.com>

1.2 Objetivos e Hipótese

Um problema comum nos Sistemas de Recomendação está relacionado à qualidade dos itens recomendados. Usualmente, esses sistemas podem recomendar itens tanto relevantes quanto irrelevantes para o usuário, dependendo do seu perfil e dos itens avaliados anteriormente. Assim, **objetiva-se** que as recomendações sejam cada vez mais adequadas aos interesses dos usuários (Khan et al., 2004).

A literatura demonstra que a busca que utiliza conceitos organizados em uma base de conhecimento permite o acesso aos documentos por meio do seu significado, melhorando a qualidade e a relevância da informação recuperada. As bases de conhecimento são modeladas por meio de estruturas conceituais que se encarregam de capturar os conceitos relativos a um determinado domínio, bem como as relações entre eles (Bhogal et al., 2007). Essas estruturas conceituais podem ser modeladas por meio de uma ontologia.

Uma forma de melhorar a organização dos relacionamentos entre os conceitos que compõem uma ontologia tradicional (crisp)⁴ é por meio do uso da lógica *fuzzy*, devido à capacidade de inferir informações imprecisas (Morrissey, 1990), indicando a intensidade com a qual um conceito está relacionado ao outro (Yaguinuma et al., 2010).

Dessa forma, o **objetivo principal** deste trabalho é propor uma abordagem que permita utilizar essas estruturas conceituais que aplicam conceitos da lógica *fuzzy* para um determinado domínio, com o intuito de melhorar a precisão dos resultados nos processos de recomendação.

Na literatura e na Web, existem diversas ontologias crisp amplamente reconhecidas e validadas em diferentes domínios. Existem também diversas metodologias que permitem a criação de ontologias desde o início. No entanto, existem poucas pesquisas que visam construir uma ontologia *fuzzy* ou a “fuzzificar” uma ontologia crisp já existente. Assim, um **segundo objetivo** é propor uma abordagem que permita converter as ontologias crisp existentes em ontologias *fuzzy*, as quais permitem representar os graus de similaridade entre os conceitos.

Os sistemas de recomendação usam os perfis de usuário como base para a recomendação. Esses perfis podem ser representados por uma lista de itens associada ao usuário, por um perfil que represente o histórico de buscas rea-

⁴Na teoria clássica, os conjuntos são denominados "crisp" e um dado elemento do universo em discurso (domínio) pertence ou não pertence ao referido conjunto.

lizadas ou por uma ontologia que estruture os itens buscados pelo usuário (Ricci et al., 2011). Um desafio dos Sistemas de Recomendação é determinar quais itens são relevantes quando uma nova busca é efetuada. Assim, na literatura são propostos alguns métodos que permitem propagar o interesse do usuário para um determinado item, considerando os itens similares associados Jiang e Tan (2006); He e Fang (2008); Mohammed et al. (2010). Geralmente, essas relações de similaridade são estabelecidas pelo mecanismo de recomendação durante a busca que calcula uma matriz de similaridade a cada nova busca. Deve-se considerar também que o interesse associado a um determinado item está limitado pela estrutura que representa o perfil do usuário. Sendo que, a propagação de interesses pode ser feita unicamente entre nós irmãos ou pais.

Dessa forma, o **terceiro objetivo** deste trabalho é elaborar uma ontologia de preferência do usuário que permita armazenar e atualizar os interesses para cada usuário e que, por meio de um mecanismo de propagação, permita utilizar os relacionamentos de similaridade estabelecidos na ontologia *fuzzy* de domínio.

Além disso, o **quarto objetivo** é avaliar se o uso de uma ontologia *fuzzy* de domínio melhora a precisão dos resultados em uma aplicação de recomendação. Para isso, os resultados obtidos a partir de um Sistema de Recomendação apoiado na abordagem proposta neste trabalho são comparados aos resultados de um Sistema de Recomendação apoiado em uma abordagem baseada em agrupamento. Tendo em vista que as técnicas de agrupamento permitem obter uma representação taxonômica dos itens similar à organização de uma ontologia torna a comparação dos resultados bastante natural.

O desenvolvimento deste trabalho se pautou na **hipótese** que o uso de uma ontologia *fuzzy* de domínio associada a uma ontologia de preferência do usuário no processo de recomendação permite uma maior precisão dos itens recomendados.

1.3 Principais Resultados

De forma a cumprir os objetivos propostos neste projeto de mestrado, inicialmente, foi feito um estudo sobre ontologias e lógica *fuzzy*. Além disso, foram estudadas as metodologias que permitem a construção de ontologias *fuzzy* e os métodos de propagação de interesse do usuário. Esse estudo permitiu propor uma abordagem para a fuzzificação de ontologias e para a propa-

gação do interesse do usuário em Sistemas de Recomendação. As principais contribuições e os resultados alcançados com este projeto foram:

- A proposta de uma abordagem de recomendação baseada em ontologia *fuzzy* de domínio e em ontologia de preferência do usuário (ObRA - Ontology-based Recommendation Approach);
- A adaptação de uma metodologia de construção de ontologias (UPON Unified Process for ONtology building) para permitir a construção ou fuzzificação de uma ontologia crisp (UPFON - Unified Process for Fuzzy ONtology building);
- A fuzzificação de uma ontologia crisp de dispositivos móveis. Para isso, foram estudadas e selecionadas as principais características dos dispositivos móveis apresentados no *Wireless Universal Resource FiLe* (WURFL⁵);
- A proposta e a elaboração de um método que permita propagar o interesse do usuário considerando a similaridade dos itens estabelecidos na ontologia *fuzzy*.

1.4 Organização do Trabalho

Esta dissertação está organizada em outros cinco capítulos, além deste introdutório. Na Figura 1.1, é descrita a organização geral do trabalho. Os conteúdos de cada capítulo são detalhados a seguir.

No Capítulo 2 é apresentado o referencial teórico sobre Mineração de Textos, Lógica Fuzzy, Ontologias e Sistemas de Recomendação;

No Capítulo 3 é apresentada uma revisão da literatura sobre Sistemas de Recomendação, ontologias de domínio e ontologias fuzzy e uma discussão dos principais conceitos e técnicas envolvidas neste trabalho;

No Capítulo 4 é descrita a abordagem ObRA. Além disso, são discutidas as principais ideias em torno da abordagem e suas atividades. Nesse capítulo também é descrito um estudo de caso que permite determinar a viabilidade da utilização da abordagem proposta;

No Capítulo 5 a avaliação da abordagem é detalhada;

Por fim, no Capítulo 6 são apresentadas as conclusões do trabalho, suas contribuições e limitações, além dos trabalhos futuros.

⁵<http://wurfl.sourceforge.net/>

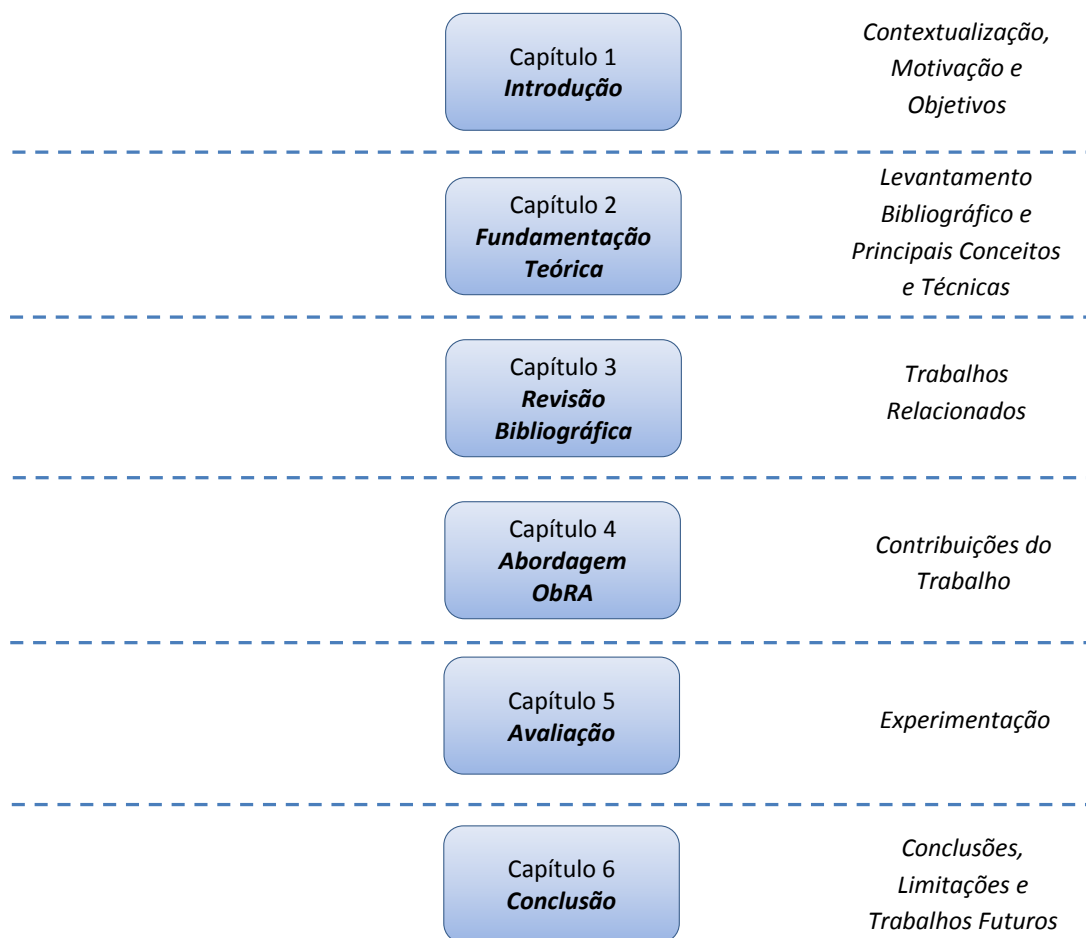


Figura 1.1: Organização da Dissertação

Fundamentação Teórica

Para poder contextualizar o problema do trabalho torna-se necessário compreender alguns termos e conceitos relacionados com o trabalho realizado nesta dissertação. Com isso, no presente capítulo são descritos os conceitos de Mineração de Textos, apresentados na Seção 2.1, focando-se principalmente no cálculo de proximidade e algoritmos de agrupamento; os conceitos referentes a lógica *fuzzy*, com foco nos graus de pertinência, são apresentados na Seção 2.2; os conceitos referentes as ontologias são apresentados na Seção 2.3, especialmente as ontologias de domínio e de usuário; e, finalmente, os conceitos e os modelos dos sistemas de recomendação são apresentados na Seção 2.4.

2.1 Mineração de Textos

A Mineração de Textos (MT) pode ser definida como um conjunto de técnicas e processos para descoberta de conhecimento inovador a partir de dados textuais (Ebecken et al., 2005). Em um contexto no qual grande parte da informação corporativa como e-mails, memorandos e blogs industriais são registrados em linguagem natural, a Mineração de Textos surge como uma poderosa ferramenta para a gestão do conhecimento. Nesse sentido, o objetivo da prática de Mineração de Textos centra-se na busca por padrões, tendências e regularidades em documentos escritos em linguagem natural.

A Mineração de Textos foi inspirada na Mineração de Dados e por vezes é

considerada uma subárea da mesma. A Mineração de Dados pode ser definida como a extração do conhecimento implícito, previamente desconhecido e potencialmente útil, ou a busca por relações e padrões globais existentes em base de dados (Frawley et al., 1992). A principal diferença entre os processos de Mineração de Textos e Mineração de Dados é que a primeira lida com dados não-estruturados, como páginas HTML ou arquivos XML. Assim, na Mineração de Textos, o primeiro desafio é obter alguma estrutura que represente os documentos textuais.

O processo de Mineração de Textos pode ser dividido em cinco grandes etapas, formando um ciclo no qual, ao final, obtém-se o conhecimento acerca dos dados analisados. As etapas são: identificação do problema, pré-processamento, extração de padrões, pós-processamento e utilização do conhecimento, como pode ser observado na Figura 2.1.

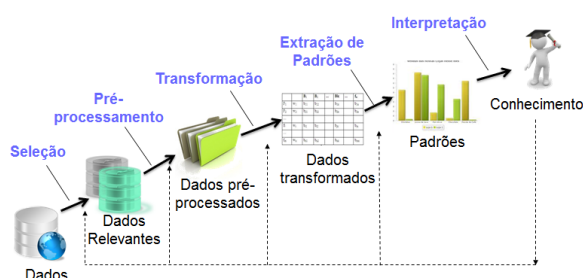


Figura 2.1: Etapas do processo de Mineração de Textos (Rezende, 2003)

Nas próximas seções serão detalhadas as etapas que fazem parte da Mineração de Textos, com foco principalmente na etapa de pré-processamento e extração de padrões.

2.1.1 Identificação do Problema

Esta primeira etapa é considerada uma das mais importantes etapas. Nesta etapa, o especialista do domínio identifica e delimita o problema, o subdomínio do problema, a coleção de textos a ser analisada, a existência de algum conhecimento prévio que possa ser utilizado na análise, o que se espera obter e como os resultados poderão ser utilizados. Para isso Rezende (2003) define quatro questões essenciais a serem respondidas nesta etapa:

- Quais são as principais metas do processo.
- Quais critérios são importantes para o problema.

- O conhecimento extraído deve ser compreensível a seres humanos ou um modelo do tipo caixa-preta também é apropriado.
- Qual deve ser a relação entre simplicidade e precisão do modelo extraído.

As respostas a estas questões representam decisões que guiam os próximos passos do processo e têm reflexo no desempenho da aplicação. No contexto deste trabalho, o objetivo do processo de Mineração de Textos é extrair uma matriz de similaridade entre um conjunto de documentos os quais servirão como base para estabelecer relacionamentos *fuzzy* de similaridade entre conceitos de uma ontologia de domínio. Essa matriz de similaridade permitirá também, utilizar os valores de similaridade com o objetivo de realizar um agrupamento de documentos, deixando aberta a possibilidade de realizar a comparação das duas técnicas em ambientes de aplicações de recomendação.

Um aspecto de grande importância nesta etapa é determinar a coleção textual a ser utilizada, devendo-se selecionar textos que sejam mais relevantes ao domínio de aplicação do conhecimento a ser extraído. Depois de ter o conjunto de textos é necessário convertê-los em um padrão de texto puro para que possam ser processados de forma mais fácil.

2.1.2 Pré-processamento

Nesta etapa encontra-se a principal diferença entre Mineração de Dados e Mineração de Textos. Nos dois casos o problema resume-se ao fato que os dados não se encontram em um formato adequado para a etapa de Extração de Padrões. É por isso que faz-se necessário adequar a coleção de textos para um formato manipulável por algoritmos de extração de conhecimento, além de aplicar-lhes um tratamento de limpeza, em geral, redução do volume de textos ([Batista, 2003](#)).

O objetivo do pré-processamento é extrair textos escritos em linguagem natural, inerentemente não estruturados em uma representação estruturada, concisa e manipulável. Para tal, são executadas atividades de tratamento e padronização da coleção de textos, seleção dos termos (palavras) mas significativas, e, por fim, representação da coleção textual em um formato estruturado que preserve as características necessárias aos objetivos definidos na etapa de identificação do problema ([Feldman e Sanger, 2006](#)).

Nesta etapa, a partir da coleção de textos já definida, ocorre a preparação dos textos para poder extrair os termos importantes da coleção. Estes termos

devem ser representados na forma matriz atributo-valor, conforme ilustrado na Figura 2.2 e detalhado a seguir.



Figura 2.2: Etapas do pré-processamento de Textos. Adaptado de Moura (2006)

Na **padronização dos textos**, geralmente, os documentos são convertidos para uma forma de texto plano sem formatação. Na **limpeza dos textos**, elimina-se dos mesmos as *stopwords*, que são aquelas palavras que nada acrescentam à representatividade dos termos ou que sozinhas nada significam, como artigos, pronomes e advérbios. Essa eliminação reduz significativamente a quantidade de termos diminuindo o custo computacional das próximas etapas (Manning et al., 2008). Posteriormente, busca-se identificar as variações morfológicas e termos sinônimos. Para tal, pode-se, por exemplo, reduzir uma palavra à sua raiz por meio de processos de *stemming* ou mesmo usar dicionários ou *thesaurus*. Além disso, é possível buscar na coleção a formação de termos compostos, ou n-gramas, que são termos formados por mais de um elemento, porém com um único significado semântico (Manning et al., 2008), (Conrado et al., 2009).

Na **extração de termos**, faz-se necessária a busca por padrões que simplifiquem as diversas formas de representação dos termos. Entre as técnicas mais utilizadas para este fim, encontram-se a radicalização, lematização e substantivação. Uma vez extraídos os termos que são os atributos que representam os documentos textuais, é obtida uma representação estruturada. Quando é feita a representação dos documentos, o modelo espaço vetorial é tipicamente usado (Salton, 1989). Nesta forma espaço vetorial dos documentos englobam-se os termos utilizados e algum **valor do atributo**. Cada documento é representado por um vetor (d_i), e cada posição desse vetor corresponde a uma dimensão (atributo/termo) do documento (t_i). Ao unir todos

os vetores que representam os documentos de uma coleção, forma-se a **matriz atributo-valor** apresentada na Figura 2.3.

	t_1	t_2	t_3	...	t_M	Classe
d_1	a_{11}	a_{12}	a_{13}	...	a_{1M}	C_1
d_2	a_{21}	a_{22}	a_{23}	...	a_{2M}	C_2
d_3	a_{31}	a_{32}	a_{33}	...	a_{3M}	C_2
\vdots	\vdots	\vdots	\vdots	\ddots	\vdots	\vdots
d_N	a_{N1}	a_{N2}	a_{N3}	...	a_{NM}	C_N

Figura 2.3: Representação da matriz atributo-valor para N documentos e M atributos

O modelo espaço vetorial é simples e permite o uso dos métodos tradicionais de extração de padrões que lidam com vetores de atributos numéricos. Por meio da matriz atributo-valor, cada documento pode ser representado como um vetor $\vec{d}_i = (a_{i1}, a_{i2}, \dots, a_{iM})$. Geralmente, o valor da medida a_{ij} é obtido de duas formas:

1. Um valor que indica se um determinado termo está ou não está presente (representação binária); e
2. Um valor que indica a importância ou distribuição do termo ao longo da coleção de documentos, por exemplo o valor da frequência do atributo no documento (*tf - term frequency*). Outras formas, baseadas em critérios de normalização ou padronização podem ser encontradas em [Salton e Buckley \(1988\)](#) e [Liu et al. \(2005\)](#). Entre elas pode-se considerar a quantidade de documentos em que o atributo aparece (*df - document frequency*), a frequência inversa do atributo na coleção de documentos (*idf - inverse document frequency*), e a combinação da frequência no documento com a frequência inversa na coleção de documentos (*tf-idf*).

A abordagem *bag-of-words* é uma das formas de representação de documentos textuais baseada no modelo espaço vetorial e uma das mais utilizadas na Mineração de Textos. Cada palavra encontrada na coleção pode tornar-se um atributo nesta representação ([Salton, 1989](#)).

Deve-se ressaltar, que esta etapa de Pré-processamento pode ser redefinida e então repetida após as próximas etapas, uma vez que a descoberta de alguns padrões pode levar a estabelecer melhorias a serem empregadas sobre o valor do atributo utilizado na matriz atributo-valor.

2.1.3 Extração de Padrões usando Agrupamento de Dados

Nesta etapa são definidas as tarefas a serem realizadas de acordo com o objetivo do processo. Estas tarefas podem ser preditivas ou descritivas (Conrado et al., 2009).

- **Tarefas Preditivas:** consistem na generalização de exemplos ou experiências passadas com respostas conhecidas. Essas tarefas aplicam os chamados modelos de aprendizado de máquina supervisionado, devido a que as categorias são sempre pré-conhecidas e disponíveis junto aos dados. Esses modelos podem ser divididos em tarefas de classificação ou regressão.
- **Tarefas Descritivas:** consistem na identificação de comportamentos intrínsecos da coleção de textos, sendo que esses dados são exemplos não rotulados. Nessas tarefas são usados modelos de aprendizado de máquina não-supervisionado, e as principais tarefas são regras de associação, agrupamento de dados (*clustering*), sumarização e visualização.

Em tarefas de agrupamento, o objetivo é organizar um conjunto de objetos em grupos, baseados em uma medida de proximidade, na qual objetos de um mesmo grupo são altamente similares entre si, mas dissimilares em relação aos objetos de outros grupos (Everitt et al., 2001). Em outras palavras, o agrupamento é baseado no princípio de maximizar a similaridade interna dos grupos e minimizar a similaridade externa dos grupos. Este método de agrupamento, pertence aos métodos de aprendizado de máquina não-supervisionados. A diferença deste método com os algoritmos de classificação é que este não possui classes ou rótulos predefinidos para treinamento de um modelo (Jain et al., 1999), (Han e Kamber, 2006).

O processo de agrupamento depende de dois fatores principais: (1) uma medida de proximidade e (2) uma estratégia de agrupamento. As **medidas de proximidade** determinam como a similaridade entre dois objetos é calculada. Sua escolha influencia a forma como os grupos são obtidos e depende dos tipos de variáveis ou atributos que representam os objetos. As **estratégias de agrupamento** são métodos e algoritmos para definição dos grupos. Em geral, pode-se classificar os algoritmos de agrupamento em métodos particionais e métodos hierárquicos.

Medidas de Proximidade

A escolha da medida de proximidade para calcular quão similares são dois objetos é fundamental para aplicar técnicas de classificação, agrupamento, entre outros. Essa escolha depende das características do conjunto de dados, principalmente dos tipos e escala dos dados. As medidas de proximidade podem calcular tanto a similaridade quanto a dissimilaridade (ou distância) entre objetos. A seguir, é descrita a medida de similaridade comumente utilizada em dados textuais: *Cosseno*. Para tal, considere dois documentos $x_i = (x_{i1}, x_{i2}, \dots, x_{im})$ e $x_j = (x_{j1}, x_{j2}, \dots, x_{jm})$, representados no espaço vetorial m -dimensional, no qual cada termo da coleção representa uma dessas dimensões.

A medida de similaridade Cosseno é definida de acordo com o ângulo cosseno formado entre os vetores de dois documentos, conforme mostrado na Equação 2.1 (Tan et al., 2005; Feldman e Sanger, 2006).

$$\text{cosseno}(x_i, x_j) = \frac{x_i \cdot x_j}{|x_i||x_j|} = \frac{\sum_{l=1}^m x_{il}x_{jl}}{\sqrt{\sum_{l=1}^m x_{il}^2} \sqrt{\sum_{l=1}^m x_{jl}^2}} \quad (2.1)$$

O valor da medida está no intervalo $[0, 1]$. Assim, se o valor da medida de similaridade Cosseno é 0, o ângulo entre x_i e x_j é 90° , ou seja, os documentos não compartilham nenhum termo. Por outro lado, se o valor da similaridade for próximo a 1, o ângulo entre x_i e x_j é próximo de 0° , indicando que os documentos compartilham termos e são similares.

Pode-se encontrar uma variedade de medidas de proximidades na literatura. Nessa seção, foi apresentada a medida de cosseno que está relacionada com este projeto. Uma revisão mais extensa está disponível nos trabalhos de Everitt et al. (2001) e Tan et al. (2005).

Estratégias de Agrupamento

Os métodos de agrupamento podem ser classificados considerando diferentes aspectos. Em geral, as estratégias de agrupamento podem ser organizadas em dois tipos: (1) agrupamento particional e (2) agrupamento hierárquico. No agrupamento particional a coleção de documentos é dividida em uma partição simples de k grupos, como ilustrado na Figura 2.4. Enquanto no agrupamento hierárquico é produzido uma sequência de partições aninhadas, ou seja, a coleção textual é organizada em grupos e subgrupos de documentos, como ilustrado na Figura 2.5 (Feldman e Sanger, 2006).

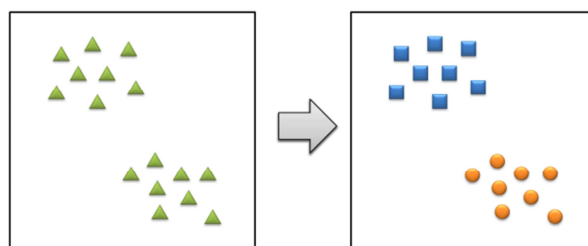


Figura 2.4: Exemplo de agrupamento de documentos particional (Rossi, 2011)

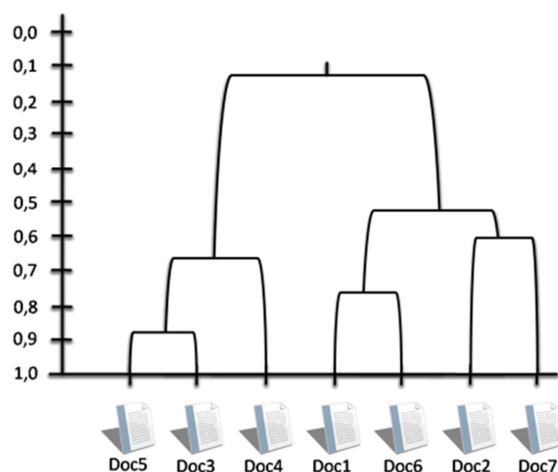


Figura 2.5: Exemplo de um dendrograma obtido por um agrupamento de documentos hierárquico (Rossi, 2011)

Segundo Zhao e Karypis (2002), as hierarquias de grupos obtidas utilizando algoritmos de agrupamento hierárquicos provêm uma visão dos documentos em diferentes níveis de granularidade, o que facilita a visualização e a interação ao explorar uma grande quantidade de documentos.

Os algoritmos de agrupamento hierárquicos podem ser divididos em aglomerativos ou divisivos. Nos algoritmos de agrupamento hierárquico divisivos, inicialmente todos os objetos estão contidos em um único grupo, e então, são realizadas sucessivas divisões nas partições até que se tenha um objeto por grupo ou até que um critério de parada seja atingido. Já os algoritmos de agrupamento hierárquicos aglomerativos, inicialmente cada objeto corresponde a um grupo. A partir disso, são aglomerados pares de grupos conforme algum critério de parada seja atingido. O resultado de um algoritmo de agrupamento hierárquico aglomerativo é uma representação denominada *dendrograma*, como a apresentada na Figura 2.5. Essa representação mostra os grupos e subgrupos obtidos pelo algoritmo de agrupamento, e a similaridade entre os grupos da hierarquia (Rossi, 2011).

A diferença principal entre os algoritmos de agrupamento hierárquico aglomerativo está no critério de seleção do par de grupos mais próximo. Os três critérios mais conhecidos são:

- *Single-Link* (Everitt et al., 2001; Sneath, 1957): utiliza o critério de vizinho mais próximo, no qual a distância entre dois grupos é determinada pela distância do par de documentos mais próximos, sendo cada documento pertencente a um desses grupos.
- *Complete-link* (Everitt et al., 2001; Sorensen, 1948): utiliza o critério de vizinho mais distante, ao contrário do *Single-Link*, e a distância entre dois grupos é a maior distância entre um par de documentos, sendo cada documento pertencente a um grupo distinto.
- *Average-Link* (Everitt et al., 2001; Sokal e Michener, 1958): a distância entre dois grupos é definida como a média das distâncias entre todos os pares de documentos em cada grupo, cada par é composto por um documento de cada grupo.

2.1.4 Pós-processamento e Utilização do conhecimento

O Pós-processamento é a etapa de validação das descobertas efetuadas. É a etapa de avaliação do conhecimento obtido e a apresentação do mesmo, seja por ferramentas de visualização ou simplesmente por tabelas de resultados. Nesta etapa o especialista do domínio e o de Mineração de Textos devem trabalhar juntos, procurando responder a questões como: representatividade do conhecimento obtido; o que há de novo nos resultados encontrados; de que maneira o conhecimento do especialista difere do obtido; validação dos resultados obtidos; e de que maneira os resultados obtidos devem ser utilizados.

2.2 Lógica Fuzzy

A lógica nebulosa, também conhecida como lógica *fuzzy*, surgiu por volta de 1965, com os estudos do professor Lotfi A. Zadeh (Zadeh, 1965). A característica dessa teoria é prover um método de tradução de expressões verbais, vagas, imprecisas e/ou qualitativas, em valores numéricos.

Na lógica *fuzzy*, o valor de verdade de uma proposição pode ser um subconjunto *fuzzy* de qualquer conjunto parcialmente ordenado, ao contrário dos sistemas lógicos binários, nos quais o valor de verdade só pode assumir os

valor de verdadeiro (1) ou falso (0). Na lógica *fuzzy*, os valores verdade são expressos linguisticamente, por exemplo, frio, muito frio, quente, muito quente. Cada termo linguístico pode ser interpretado como um subconjunto *fuzzy* do intervalo unitário.

Na teoria de conjuntos clássica, um elemento pertence ou não a um dado conjunto. Dado um universo U e em elemento particular $x \in U$, o grau de pertinência $\mu_A(x)$ com respeito a um conjunto $A \subseteq U$ é dado por:

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

A função $\mu_A(x) : U \rightarrow 0,1$ é chamada de função característica na teoria clássica de conjuntos.

2.2.1 Noções de lógica *Fuzzy*

A lógica *fuzzy* difere da lógica clássica em seu mapeamento de verdade e falsidade. Nessa lógica o valor de verdade de uma proposição pode ser um subconjunto *fuzzy*, por exemplo, baixo, médio e alto, de qualquer conjunto parcialmente ordenado, como ilustrado na Figura 2.6.

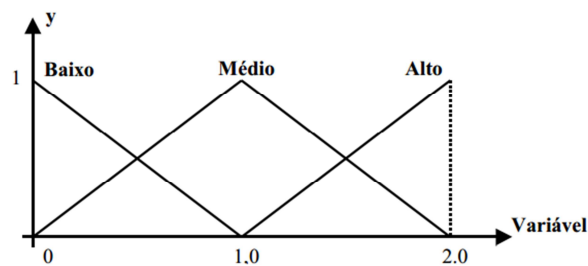


Figura 2.6: Domínio das variáveis linguísticas

Por outro lado, como ilustrado na Figura 2.7, nos sistemas lógicos clássicos o valor verdade só pode assumir dois valores: verdadeiro(1) ou falso(0).

Por exemplo, ao se dizer que uma determinada pessoa é alta, é perfeitamente entendido, mesmo que não se defina exatamente a partir de que altura uma pessoa pode ser considerada alta (Shaw e Simões, 1998). Sem se estabelecer uma referência exata, é impossível determinar se a afirmação “João é alto” é verdadeira ou falsa com a lógica binária. Se a referência fosse 1.8m, uma pessoa com 1.81m seria considerada “alta”, enquanto outra, de 1.79m seria “não-alta”. Na Figura 2.8 observa-se uma linha que divide as duas classes. Essa linha possui uma mudança brusca, e um valor binário 0 ou 1 indica

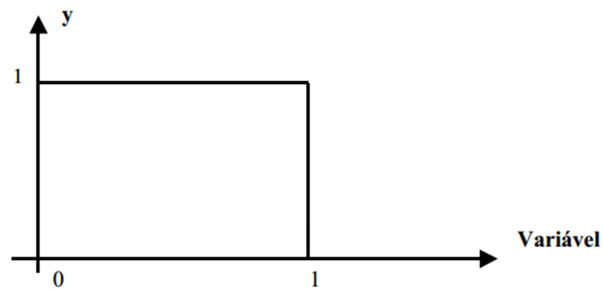


Figura 2.7: Domínio da lógica clássica

em qual classificação se encontraria uma pessoa em termos de sua altura em metros.

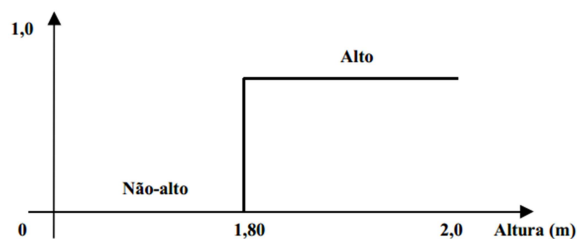


Figura 2.8: Classificação Binária entre ser alto ou não-alto (Shaw e Simões, 1998)

Cada uma dessas afirmações linguísticas representa uma verdade parcial, um certo grau de verdade, ou um certo grau de pertinência a um conjunto. Na lógica *fuzzy*, um elemento pode pertencer de forma parcial a um conjunto, com um certo grau, por exemplo 80% ou 25%. Na lógica clássica, o grau de pertinência é sempre 0% ou 100% (Shaw e Simões, 1998).

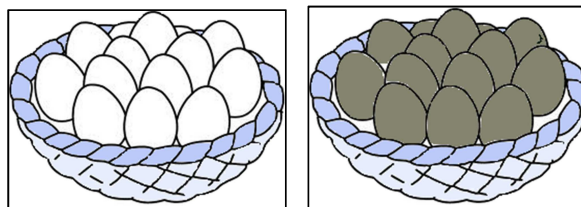


Figura 2.9: Cesto com bolinhas brancas e escuras adaptada de Pinho (1999)

Pode-se observar uma descrição mais simples da lógica *fuzzy* na Figura 2.9. A pergunta relativa com a figura é: qual é a cor das bolinhas contidas no cesto? Nesta Figura pode-se distinguir claramente que no cesto da esquerda há bolinhas brancas e no da direita bolinhas escuras. Se agora analisamos a

Figura 2.10, nota-se uma dificuldade antes inexistente na tentativa de classificar a cor do conteúdo do cesto.

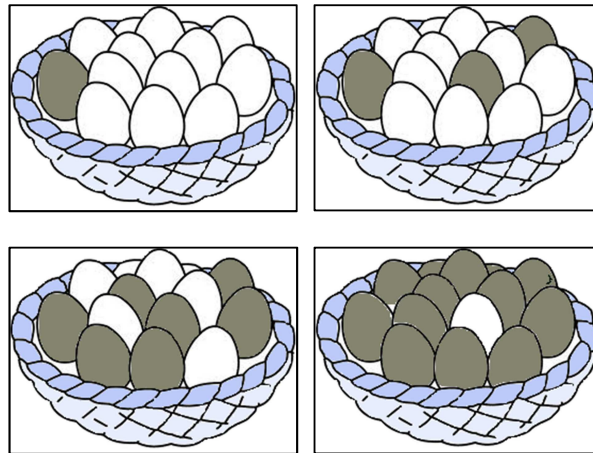


Figura 2.10: Cesto com bolinhas variadas (Pinho, 1999)

Provavelmente a resposta mais adequada à pergunta ao se analisar a Figura 2.10 seja: quase totalmente branca; grande parte branca; grande parte escura; quase totalmente escura. Nessas respostas nota-se a criação imediata de uma escala de valores, responsável por quantificar o conteúdo dos cestos. Sob a ótica da matemática convencional, seria impossível descrever a cor do conteúdo do cesto da Figura 2.10 como “grande parte branca”. Porém a lógica *fuzzy*, por meio do uso de variáveis linguísticas, consegue captar a incerteza associada a essas variáveis e traduzi-la para o modelamento matemático (Jané, 2004).

Formalmente, um conjunto *fuzzy* A é definido por uma função de pertinência $\mu_A : [0 - 1]$. Essa função associa a cada elemento x do conjunto *fuzzy* A um grau $\mu_A(x)$ de pertinência, isto representa o grau de compatibilidade entre x e o conceito expresso por A (Sandri e Correa, 1999):

- $\mu_A(x) = 1$ indica que x é completamente compatível com A ;
- $\mu_A(x) = 0$ indica que x é completamente incompatível com A ;
- $\mu_A(x) > 0$ e $\mu_A(x) < 1$ indica que x é parcialmente compatível com A , com grau $\mu_A(x)$

A forma de expressar o conhecimento é tipicamente feito com regras do tipo condição-ação. Supondo que uma variável de entrada (condição) “salário” (S) esteja associada a uma variável de saída (ação) chamada “motivação pessoal” (M), uma regra típica é dada por:

$$SE(SBaixo)ENTAO(Mpouca)$$

A ideia geral é representar o conhecimento por um conjunto de regras nas quais as condições são dadas a partir de um conjunto de termos linguísticos associados às variáveis de entrada/saída do processo. As regras do tipo Se - Então são chamadas de regras *fuzzy* (Pagliosa, 2006). Uma regra *fuzzy* pode ser:

$$SE(xa_i)E(yb_i)e...ENTAO(zc_i).$$

onde as variáveis de entrada (x, y, \dots) são referidas como antecedentes e a implicação lógica da regra está na formulação de uma conexão entre causa e efeito, ou uma condição e sua consequência.

2.2.2 Funções de Pertinência

As funções de pertinência são de diferentes formas, dependendo do conceito que se deseja representar e do contexto onde estão inseridas (Gonçalves, 2007).

Considere a variável linguística “salário”, constituída dos seguintes termos: $T(\text{salario}) = \text{baixo, medio, alto}$, representados pelos conjunto *fuzzy* B, M e A , respetivamente, definidos por suas funções de pertinência. Observa-se que na Figura 2.11:

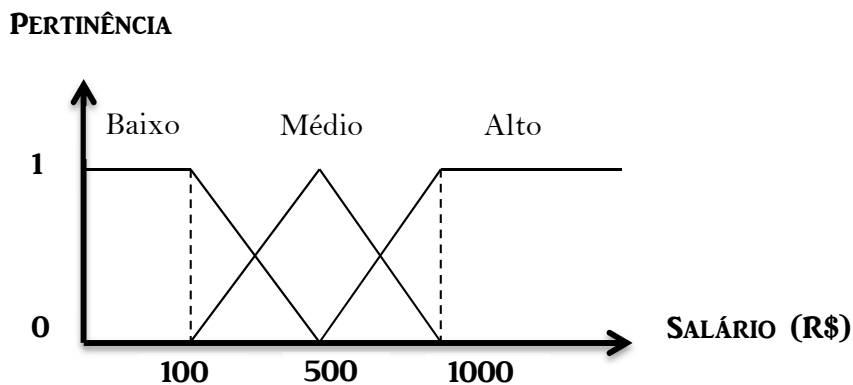


Figura 2.11: Funções de pertinência para a variável T. Adaptado de Costa et al. (2007)

- Pessoas que ganham até R\$100 apresentam um grau de pertinência igual

a 1 no conjunto B; o grau de pertinência nesse conjunto decresce à medida que o salário aumenta;

- Uma pessoa que ganha R\$500 é totalmente pertencente ao conjunto M;
- Pessoas que ganham acima de R\$700 (aproximadamente), apresentam grau de pertinência diferente de 0 em A.
- Pessoas que ganham acima de R\$1000, definitivamente estão com um salário alto.

Em alguns casos, a teoria clássica é satisfatoriamente utilizada, entretanto para alguns outros problemas pode não ser suficiente, como é o caso dos conjuntos cujo limite de pertinência e não pertinência não são claros, e existe uma transição entre esses dois grupos. Por exemplo:

- O conjunto de pessoas altas
- O conjunto de carros velozes
- O conjunto de pessoas com alta renda

Utilizando os conjuntos *fuzzy* pode-se definir critérios e graus de pertinência para cada uma dessas situações, como ilustrado na Figura 2.12.

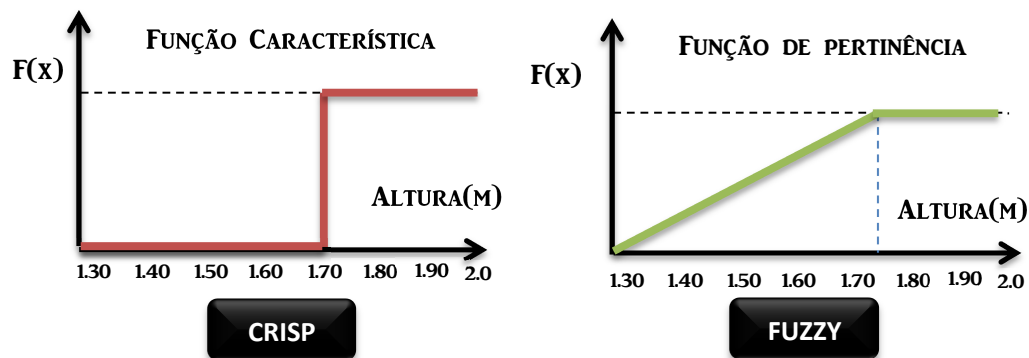


Figura 2.12: Comparação entre função característica e função de pertinência para o conjunto de pessoas altas (Moraes e Marley, 2007)

2.3 Ontologias

O termo ontologia é originário da filosofia. Ontologia é um ramo da filosofia que lida com a natureza e organização do ser. No contexto da pesquisa sobre

ontologias, filósofos tentam responder as questões “O que é um ser?” e “Quais são as características comuns de todos os seres?” (Maedche e Staab, 2001).

Há diversas definições de ontologias na literatura que consideram diferentes pontos de vista sendo até mesmo complementares para uma mesma realidade. Entre as várias definições, Fensel (2000) define uma ontologia como uma especificação formal explícita de uma conceitualização compartilhada. Nessa definição, pode-se constatar algumas características importantes que a ontologia deve possuir (a ontologia deve ser explícita, formal, e descrever um conhecimento comum a um grupo), porém não se descreve como é a estrutura de uma ontologia.

Segundo Gómez-Pérez (1999), uma ontologia é um conjunto de termos ordenados hierarquicamente para descrever um domínio que pode ser usado como um esqueleto para uma base de conhecimentos. Segundo Guimarães (2002), uma ontologia provê uma estrutura básica na qual logo acima pode-se construir uma base de conhecimentos. A ontologia fornece um conjunto de conceitos e termos para descrever um determinado domínio, enquanto a base de conhecimento usa esses termos para descrever uma determinada realidade.

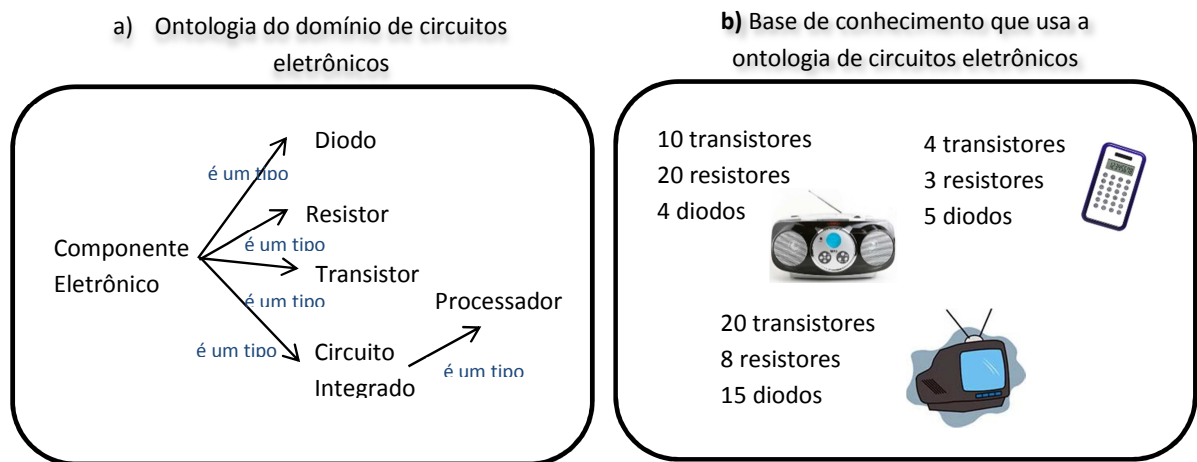


Figura 2.13: Comparação entre ontologia e base de conhecimento. Adaptado de Guimarães (2002)

Para esclarecer a diferença entre ontologia e base de conhecimentos, na Figura 2.13, baseada do trabalho de Guimarães (2002), apresenta-se um exemplo ilustrativo. Nesse exemplo, pode-se definir uma ontologia para o domínio de circuitos eletrônicos. A ontologia resultante incluiria conceitos específicos como circuito integrado, resistor, transistor, entre outros; incluiria também relações entre esses conceitos, por exemplo, o processador é um tipo de cir-

cuito integrado. Na hora de colocar na base de conhecimentos um circuito de um rádio, ou seja, descrever a realidade de um rádio, esse pode estar formado por 10 transistores, 20 resistores e 4 diodos. Pode-se observar que apesar da realidade descrita na base de conhecimento variar para o rádio e para outros dispositivos (mudando a quantidade de elementos que formam um equipamento), a ontologia empregada continua sendo a mesma, ou seja, usam-se os mesmos conceitos para descrever as duas realidades.

Já [Guarino \(1997\)](#) define a uma ontologia como uma conceitualização compartilhada de um determinado domínio. Ela é composta de um conjunto de conceitos dentro desse domínio, sendo esses organizados como uma taxonomia, e de relações entre esses conceitos. A ontologia possui também axiomas, ou seja, regras pertinentes ao domínio em questão.

2.3.1 Tipos de Ontologias

[Gómez-Pérez \(1999\)](#) define diferentes tipos de ontologia. De acordo com o grau de genericidade dessas ontologias tem-se:

- Ontologias de Representação: definem as primitivas de representação, como *frames*, axiomas, atributos, etc., de forma declarativa. Essa ideia abstrai os formalismos de representação, porém traz desvantagens ([Chandrasekaram e Josephson, 1997](#)).
- Ontologias gerais (ou de topo): trazem definições abstratas necessárias para a compreensão de aspectos do mundo, como tempo, processos, papéis, espaço, entre outros.
- Ontologias centrais (*core ontologies*) ou genéricas de domínio: definem os ramos de estudo de uma área e/ou conceitos mais genéricos e abstratos dessa área. Por exemplo, uma ontologia de direito inclui conhecimentos normativos, de agências legais, comportamentos permitidos, etc.
- Ontologias de domínio: tratam de um domínio mais específico de uma área genérica de conhecimento, como direito tributário, microbiologia, etc.
- Ontologias de aplicação: procuram solucionar um problema específico de um domínio, como identificar doenças do coração, a partir de uma ontologia de domínio de cardiologia.

Existe outra divisão quanto à classificação das ontologias, aplicável apenas para os dois últimos tipos de ontologias citados previamente:

- Ontologias de tarefas: descrevem tarefas de um domínio - como processos, planos, metas, escalonamentos, entre outros -, com uma visão mais funcional, embora declarativa, de um domínio.
- Ontologias de domínio propriamente ditas: tem uma visão mais epistemológica do domínio, focando nos conceitos e objetos do universo de discurso.

Baseados nos estudos de [Gómez-Pérez \(1999\)](#), [Fernández-López \(1999\)](#) e [Almeida e Bax \(2003\)](#), mostra-se, na Tabela 2.1, uma breve descrição dos tipos de ontologia apresentados.

2.3.2 Metodologias para à construção de Ontologias

Na literatura, muitas metodologias têm sido desenvolvidas com o intuito de sistematizar a construção e manipulação de ontologias. Existem metodologias para a construção de ontologias, para a construção de ontologias em grupo, para aprendizado sobre a estrutura de ontologias e para a integração de ontologias. Na Tabela 2.2 mostra-se uma breve descrição de algumas metodologias para construir ontologias. Nesta seção é apresentada em mais detalhes a metodologia UPON proposta por [De Nicola et al. \(2009\)](#), que foi adaptada para atingir as especificações deste trabalho.

Metodologia UPON: Unified Process for Ontology

No trabalho de DeNicola2009 é apresentada a metodologia Unified Process for Ontology Building (UPON), uma metodologia incremental para a construção de ontologias que utiliza as vantagens do Processo Unificado - Unified Process (UP) e da Linguagem de Modelagem Unificada - Unified Modeling Language (UML). Segundo DeNicola, o que distingue a metodologia UPON das outras metodologias de construção de ontologias é que além de ser de natureza iterativa, repetindo ciclicamente cada atividade do processo, é também de natureza incremental, detalhando e expandindo cada uma dessas atividades. A metodologia UPON consiste de ciclos, fases, iterações e fluxos de trabalho. Cada ciclo está dividido em quatro fases: (i) concepção, (ii) elaboração, (iii) construção, e (iv) transição. Cada fase está dividida em cinco fluxos: (i) fluxo de requisitos,

Tabela 2.1: Tipos de Ontologia (Almeida e Bax, 2003)

Abordagem	Classificação	Descrição
Quanto à função. (Mizoguchi et al., 1994)	Ontologias de domínio	Reutilizáveis no domínio, fornecem vocabulário sobre conceitos, seus relacionamentos, sobre atividades e regras que os governam.
	Ontologias de tarefa	Fornecem um vocabulário sistematizado de termos, especificando tarefas que podem ou não estar no mesmo domínio.
	Ontologias gerais	Incluem um vocabulário relacionado a coisas, eventos, tempo, espaço, casualidade, comportamento, funções, etc.
Quanto ao grau de formalismo (Uschold e Gruninger, 1996)	Ontologias altamente informais	Expressa livremente em linguagem natural.
	Ontologias semi-informais	Expressa em linguagem natural de forma restrita e estruturada.
	Ontologias semi informais	Expressa em uma linguagem artificial definida formalmente.
	Ontologia rigorosamente formal	Os termos são definidos com semântica formal, teoremas e provas.
Quanto à aplicação (Jasper e Uschold, 1999)	Ontologias de autoria neutra	Um aplicativo é escrito em uma única língua e depois convertido para uso em diversos sistemas, reutilizando-se as informações.
	Ontologias como especificação	Cria-se uma ontologia para um domínio, a qual é usada para documentação e manutenção no desenvolvimento de softwares.
	Ontologias de acesso comum à informação	Quando o vocabulário é inacessível, a ontologia toma a informação inteligível, proporcionando conhecimento compartilhado dos termos.
Quanto à estrutura (Haav e Lubi, 2001)	Ontologias de alto nível	Descrevem conceitos relacionados a todos os elementos da ontologia (espaço, tempo, matéria, objeto, evento, ação, etc), os quais são independentes do problema ou domínio.
	Ontologias de domínio	Descrevem o vocabulário relacionado a um domínio, como, por exemplo, medicina ou automóveis.
	Ontologias de tarefa	Descrevem uma tarefa ou atividade, como por exemplo, diagnósticos ou compras, mediante inserção de termos especializados na ontologia.

(ii) fluxo de análise, (iii) fluxo de projeto, (iv) fluxo de implementação e (v) fluxo de testes. Como resultado de um ciclo completo é obtida uma nova versão da ontologia. Na Figura 4.3 é ilustrada a representação da metodologia UPON.

Na Figura 4.3, no fluxo de requisitos, as necessidades semânticas são especificadas e o conhecimento é codificado na ontologia. No fluxo de análise, todos os requisitos da ontologia identificados no fluxo de requisitos são refi-

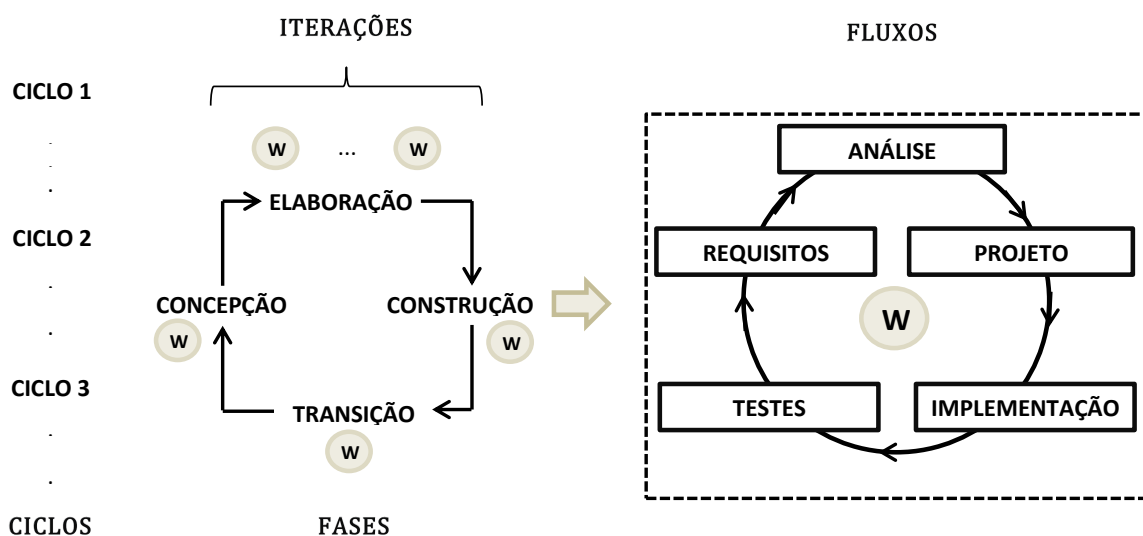


Figura 2.14: Representação da metodologia UPON de [De Nicola et al. \(2009\)](#), ressaltando os fluxos adaptados para permitir a criação ou fuzziificação de uma ontologia

nados e estruturados. No fluxo de projeto, os conceitos identificados no fluxo de análise são refinados e as relações entre eles são identificadas. No fluxo de implementação, a ontologia é expressada em termos de uma linguagem formal, por exemplo, *Web Ontology Language* (OWL) ([Antonioni e Van Harmelen, 2003](#)). Finalmente, a consistência da ontologia e da correta aplicação dos seus requisitos são verificados no fluxo de teste.

1. **Fluxo de Requisitos** O objetivo principal deste fluxo é especificar os requisitos semânticos necessários para a construção da ontologia, e estabelecer a visão do conhecimento do usuário, para assim codificar tal conhecimento na ontologia. De acordo com ([Jacobson et al., 1999](#)), essa atividade requer cooperação entre os projetistas, engenheiros do conhecimento e usuários, identificando, assim, os objetivos da ontologia do ponto de vista do usuário.

O fluxo de requisitos recebe como parâmetros de entrada as entrevistas com os especialistas do domínio e a documentação específica sobre o domínio da aplicação. Este fluxo está dividido em seis atividades:

- Determinação do domínio de interesse e do escopo, a fim de obter um melhor enfoque da realidade a ser modelada;
- Determinar o propósito para a construção da ontologia, esta atividade ajuda a obter uma melhor compreensão do domínio de inte-

resse;

- Documentar o processo por meio de histórias, que é necessário para obter uma sequência de todas as atividades que definem um cenário particular;
- Definir o léxico da aplicação, que estabelece um vocabulário com todos os termos significativos sobre um domínio específico. Por exemplo, na área do comércio eletrônico, o léxico da aplicação pode conter os termos: endereço, quantidade, cidade, pagamento devido;
- Identificar questões de competência, na qual um grupo de questões de nível conceitual são propostas. Estas questões devem ser implementadas na ontologia de forma que possam ser resolvidas utilizando o nível de inferência da ontologia. Estas questões também serão usada no fluxo de testes para testar a cobertura da ontologia;
- Identificar e priorizar os casos de uso. No contexto de ontologias, casos de uso correspondem aos caminhos do conhecimento por meio da ontologia, a ser seguidos para alcançar operações de negócios e responder as perguntas de competência.

O fluxo de requisitos tem como dados de saída um conjunto de documentos: O léxico de aplicação, as questões de competência e os modelos de caso de uso.

2. **Fluxo de Análise** O objetivo principal deste fluxo é fornecer um conjunto de atividades para estruturação e refinamento dos requisitos da ontologia identificados no fluxo de requisitos. Este fluxo recebe como parâmetros de entrada um conjunto de documentos relacionados com a área do domínio e o léxico da aplicação obtido como parâmetro de saída no fluxo de requisitos. Com esses parâmetros de entrada, o léxico da aplicação é enriquecido por meio de um léxico de domínio mais geral, para construir o léxico de referência. Este último é utilizado para produzir o glossário de referências, no qual são adicionadas definições para os termos. Este fluxo está dividido em quatro atividades:

- Aquisição dos recursos do domínio para formar um léxico de domínio. Nesta atividade, são coletados e analisados documentos de domínio, tais como artigos, relatórios técnicos, manuais, com a finalidade de permitir a elaboração de um léxico de domínio;

- Construção do léxico de referência, na qual o léxico de aplicação e o léxico de domínio são intersectados com o intuito de obter um único vocabulário de termos. Esse vocabulário é validado pela comunidade de especialistas no domínio com a ajuda dos engenheiros de conhecimento;
- Modelagem do cenário da aplicação com o uso de diagramas UML. Nesta atividade os diagramas vão representar modelos da aplicação que serão utilizadas mais tarde para a validação da ontologia. Nesses modelos cada ator/atividade deve ter um conceito correspondente na ontologia.
- Construção de um glossário referências, o qual contem todas as definições formais e informais de cada termo na ontologia. Essas definições devem estar previamente avaliadas pela comunidade especialista no domínio, podendo conter mais de uma definição para cada termo.

O fluxo de análise tem como dados de saída o léxico de domínio, o léxico de referências, os diagramas UML, e o glossário de referências.

3. **Fluxo de Projeto** O principal objetivo do fluxo do projeto é definir uma estrutura ontológica baseada no conjunto de dados presentes no glossário de referências. A estruturação dos dados é feita por meio da organização hierárquica dos termos e da definição de atributos e axiomas para cada um deles. Neste fluxo é realizado o refinamento das entidades e dos processos identificados durante o fluxo de análise, assim como a identificação dos relacionamentos existentes entre cada um deles. Este fluxo está dividido em duas atividades:

- Modelagem dos conceitos. Nesta atividade, os termos (a partir de agora referidos como conceitos) são categorizados por meio da associação de um “tipo de conceito” (e.g. ator, objeto, processo, mensagem, atributo). Esses tipos de conceitos incluem as principais categorias ontológicas definidas em reconhecidas abordagens ontológicas como (Uschold e King, 1995; Sowa, 1999; Missikoff e Taglino, 2004).
- Modelagem das hierarquias e dos relacionamentos específicos de domínio. Nesta atividade, os conceitos são hierarquicamente organizados e os relacionamentos formais entre eles são estabelecidos. A

organização hierárquica dos conceitos é feita segundo o relacionamento de generalização (relacionamento “é-um”). A taxonomia resultante pode ser estendida com outros tipos de relacionamento específico do domínio.

Uma vez realizadas as atividades do fluxo do projeto se tem como resultado uma rede estruturada de conceitos, isto é, uma rede semântica, uma ontologia.

4. **Fluxo de Implementação** O fluxo de implementação tem como objetivo realizar a representação da ontologia em uma linguagem formal (i.e. mecanismos formais para representação e especificação de linguagens, baseados na Teoria da Computação). Este fluxo está dividido em duas atividades

- Seleção da linguagem formal a ser considerada na representação da ontologia, para o qual é necessário considerar seu poder expressivo, a complexidade computacional do mecanismo de inferência que utiliza, assim como a aceitação da linguagem na comunidade científica. Como resultado de um esforço de padronização das linguagens formais na Web Semântica, o Web Ontology Language (OWL)([De Nicola et al., 2009](#)) é atualmente a linguagem mais utilizada na representação das ontologias;
- Expressar a ontologia em termos da linguagem selecionada, a ontologia modelada no fluxo de fuzziificação é traduzida à representação formal da linguagem. Como resultado do fluxo de implementação é obtida a implementação OWL da ontologia.

5. **Fluxo de Teste** O fluxo de teste tem com o objetivo avaliar a qualidade da ontologia resultante após ter seguido cada uma das atividades dos fluxos anteriores. Segundo [De Nicola et al. \(2009\)](#) a qualidade de uma ontologia é um fator multidimensional e, porém, deve ser avaliado considerando a qualidade sintática, semântica, pragmática e social da ontologia. A qualidade sintática avalia a qualidade da ontologia segundo o estilo formal adotado, quer dizer a forma em que foi representada. A qualidade semântica avalia a consistência da ontologia, verificando a ausência de conceitos contraditórios na ontologia. A qualidade pragmática se refere ao conteúdo da ontologia e à utilidade que este tem para os usuários, indiferentemente da sintaxes e semântica utilizada. A qualidade social

determinada pelo valor que tem a ontologia na comunidade (i.e. projetos com outras ontologias).

Como a qualidade sintática, tem sido considerada no momento da escolha da linguagem formal que permitiu realizar a representação da ontologia, e a qualidade social só pode ser avaliada depois da publicação da ontologia, o fluxo de testes está focado mais especificamente na avaliação da qualidade semântica e a qualidade pragmática da ontologia. Este fluxo está dividido em três atividades:

- Verificação da consistência da ontologia pode ser utilizado um *reasoner*, o qual realizara as verificações pertinentes sobre a ausência de contradições, assim como realizara a verificação do modelagem da ontologia (e.g. ausência de ciclos no modelagem da hierarquia, verificação de que as classes e as propriedades sejam disjuntas);
- Verificação a cobertura da ontologia e
- Resolução das questões de competência, estão ligadas às questões de competência e aos diagramas UML identificados em fluxos anteriores.

2.3.3 Engenharia da ontologia

A concepção de ontologias deve ser conduzida como qualquer outro projeto de software, no sentido de serem tomadas decisões de projeto que determinam sua qualidade baseada em critérios como eficiência, legibilidade, portabilidade, extensibilidade, interoperabilidade e reuso. Por isso, essa concepção deve se basear não somente em aspectos filosóficos do conhecimento acerca do domínio representado.

Existem princípios de implementação que garantem a qualidade das ontologias (Gruber, 1992). A seguir são listados os princípios mais importantes a ter em consideração na construção de ontologias (Freitas, 2006).

- Clareza: Na definição do conhecimento, deve-se ter a objetividade de definir apenas o que se presume ser útil. Definições completas, com condições necessárias e suficientes, devem ter precedência sobre definições parciais;
- Legibilidade: As definições devem guardar correspondência com as definições correntes e informais. A ontologia deve usar um vocabulário compartilhável;

Tabela 2.2: Metodologias para construção de ontologias (Almeida e Bax, 2003)

Metodologia	Descrição
Cyc	Codifica manualmente o conhecimento implícito e explícito das diferentes fontes, e, quando já se tem conhecimento suficiente na ontologia, um novo consenso pode ser obtido por ferramentas que utilizam linguagem natural (Lenat e Guha, 1990)
Uschold e King	Identifica o propósito, os conceitos e relacionamentos entre os conceitos, além dos termos utilizados para codificar a ontologia e, em seguida, documentá-la (Uschold e King, 1996)
Gruninger e Fox	Método formal que identifica cenários para uso da ontologia, utiliza questões em linguagem natural para determinação do escopo da ontologia, executa a extração sobre os principais conceitos, propriedades, relações e axiomas, definidos em Prolog (Gruninger e Fox, 1995).
Kactus	Método recursivo que consiste em uma proposta inicial para uma base de conhecimento: quando é necessária uma nova base em domínio similar, generaliza-se a primeira base em uma ontologia adaptada a ambas aplicações; quanto mais aplicações, mais genérica a ontologia (Bernaras et al., 1996)
Methontology	Constrói uma ontologia por reengenharia sobre outra utilizando-se o conhecimento do domínio; as atividades principais são especificação, conceitualização, formalização, implementação e manutenção (Fernández-López, 1999)
Sensus	Constrói ontologias a partir de outras ontologias, identificando os termos relevantes para o domínio e ligando-os à ontologia mais abrangente (Sensus, com 50 mil conceitos); um algoritmo monta a estrutura hierárquica do domínio (Swartout et al., 1997)
On-to-knowledge	Auxilia a administração de conceitos em organizações, identificando metas para as ferramentas de gestão do conhecimento e utilizando cenários e contribuições dos provedores/clientes de informação da organização (Staab et al., 2001)
Unified Process for Ontology Building	Metodologia que auxilia o processo de construção de ontologias baseada nos conceitos Processo Unificado, convertendo a construção de uma ontologia num processo iterativo incremental de atividades (De Nicola et al., 2009)

- Coerência: As inferências derivadas da ontologia definida devem ser corretas e consistentes do ponto de vista formal e informal com as definições;
- Extensibilidade: A ontologia deve permitir extensões e especializações monotonicamente e com coerência, sem a necessidade de revisão de teoria em busca de contradições;
- Mínima codificação: Devem ser especificados conceitos genéricos inde-

pendente de padrões estabelecidos para mensuração, notação e codificação, garantindo a extensibilidade. Essa genericidade é limitada pela clareza;

- **Mínimo compromisso ontológico:** Para maximizar o reuso, apenas o conhecimento essencial deve ser incluindo, gerando a menor teoria possível acerca de cada conceito, e permitindo a criação de conceitos novos, mais especializados ou estendidos.

2.4 Sistemas de Recomendação

[Herlocker et al. \(2000\)](#) definem que o objetivo de um Sistema de Recomendação é principalmente localizar os documentos que são relevantes com a informação que o usuário necessita, mas pode ser também usado para verificar a importância de uma página Web ou para descobrir os vários usos das palavras numa coleção de documentos.

Tradicionalmente, os Sistemas de Recomendação (SR) lidam com aplicações que juntam dois tipos de entidades: usuários e itens. Segundo [Resnick e Varian \(1997\)](#), os Sistemas de Recomendação são sistemas que utilizam as opiniões de uma comunidade de usuários para auxiliar indivíduos a identificarem conteúdos de interesse entre um conjunto de opções.

Segundo [Schafer et al. \(1999\)](#) e [Schafer et al. \(2001\)](#), os Sistemas de Recomendação são usados pelos sites de comércio eletrônico para sugerir produtos para seus clientes e fornecer informações que procuram auxiliar os clientes sobre qual produto deve ser adquirido. As principais funções dos sistemas de recomendação são analisar os dados dos usuários e extrair informações úteis para futuras predições ([Chen e Mccleod, 2005](#)).

Os sistemas de recomendação têm tido um importante papel no ambiente web no que diz respeito a auxiliar um usuário no acesso a informação relevante ou na escolha de itens de seu interesse. Nesse contexto itens podem ser artigos científicos ([Corrêa e Cazella, 2007](#); [Silva e Cazella, 2005](#)), filmes ([Schafer et al., 1999](#)), músicas, livros, software ou produtos de consumo em geral. Por meio de técnicas de análise de dados usadas pelos sistemas de recomendação, estes podem ajudar uma pessoa a encontrar os itens que melhor atendam suas necessidades ([Adonie et al., 2005](#)).

O desenho dos Sistemas de Recomendação depende do domínio e das características particulares dos dados disponíveis. Por exemplo, a companhia *Netflix* que tem um recomendador de filmes, frequentemente provê ranques

na escala de 1 (não-gostou) até 5 (gostou). O banco de dados que possui estas informações armazena registros sobre a qualidade das interações entre os usuários e os itens. De uma forma geral, esses sistemas buscam adquirir opiniões ou preferências sobre itens de um grupo de usuários, e usar essas opiniões para apresentar itens que possam fazer sentido para outros usuários.

Os Sistemas de Recomendação emergiram como uma área de pesquisa independente nos anos 90. O interesse nos sistemas de recomendação foi dramaticamente incrementado, como os seguintes fatos indicam (Ricci et al., 2011):

- Os Sistemas de Recomendação têm um rol importante em sites como Amazon.com, Youtube, Netflix, Yahoo, Tripadvisor, Last.fm e IMDb. Atualmente, muitas companhias estão desenvolvendo sistemas de recomendação como parte dos serviços que brindam para seus subscritores. Por exemplo, a companhia Netflix, que aluga e vende filmes online, premiaria com um milhão de dólares a equipe que conseguisse melhorar substancialmente o desempenho do seu sistema de recomendação (Koren et al., 2009).
- Existem conferências e eventos direcionados à sistemas de recomendação. Por exemplo, *ACM Recommender Systems (RecSys)*, *ACM Special Interest Group on Information Retrieval (SIGIR)*, entre outros.
- As universidades agora fornecem disciplinas, de graduação e de pós-graduação, dedicadas exclusivamente à Sistemas de Recomendação ou técnicas de recomendação.

2.4.1 Processo de Recomendação

Na Figura 2.15 é ilustrado o processo de recomendação de um SR. Nessa figura podem-se identificar duas fontes de informação necessárias como entrada do processo: (1) dados dos usuários e (2) informações a respeito dos itens e dos usuários. Idealmente esses dados relacionados ao usuário deveriam ser providos pelo próprio usuário. No entanto, essas informações também podem ser extraídas de outras fontes como navegação pelas páginas web, itens consumidos ou comprados, ranque associado a cada item, etc.

O modelo formal de um sistema de recomendação consiste em três itens, um conjunto C com os usuários, um conjunto S com os itens possíveis de recomendação, como livros, filmes ou restaurantes e uma função de utilidade

u . O espaço de itens possíveis S pode ser muito grande, com todos os itens que poderão ser recomendados. Da mesma forma, o espaço dos usuários também pode ser muito grande. A parte fundamental do modelo formal é a função de utilidade, que mede a utilidade de um item c para um usuário s , isto é, $u : C \times S \rightarrow R$. R é um conjunto ordenado, podendo conter inteiros ou números reais em um determinado intervalo (Souza, 2011).

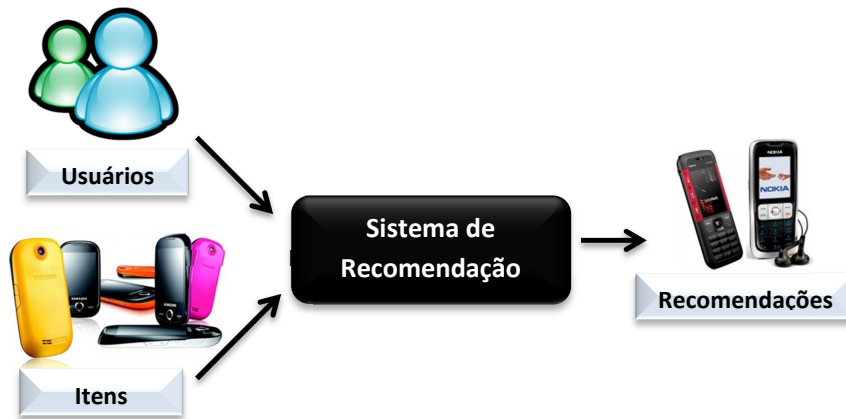


Figura 2.15: Processo de Recomendação

Usando o modelo descrito, o problema de recomendação é reduzido a escolher um item $s_i \in S$ para um usuário $c_j \in C$ que maximiza a função de utilidade. Em sistemas de recomendação, a utilidade de um item é geralmente representada por uma avaliação indicando o quanto um usuário em particular gosta de um item em particular.

Cada usuário do espaço C pode ser definido usando um perfil. Esse perfil de usuário contém informações sobre os gostos do usuário para o domínio de recomendação. A representação desse perfil depende da abordagem e do domínio. Ela pode incluir características de vários usuários, como idade, sexo, renda, estado civil, entre outros. No caso mais simples, o perfil de usuário pode conter apenas um único elemento, como o ID de usuário (Souza, 2011).

Segundo Adomavicius et al. (2005), o processo de recomendação inicia com a especificação de um conjunto inicial de pontuações que são fornecidas pelos usuários de forma explícita, ou de forma implícita inferida pelo Sistema de Recomendação. Uma vez que esta pontuação é fornecida, o sistema procura estimar a pontuação pela função R para o par usuário-item, nos itens ainda não pontuados pelo usuário alvo, como mostrado na Equação 2.2.

$$R : \text{Usuarios} \times \text{Itens} \Rightarrow \text{Pontuacoes} \quad (2.2)$$

Conceitualmente, uma vez que a função R é estimada para todo domínio de $Usuarios \times Itens$, um Sistema de Recomendação pode selecionar o item s_i com a maior pontuação para o usuário c e recomendar esses itens para o usuário alvo, como mostrado na Equação 2.3 (Adomavicius et al., 2005).

$$\forall c \in Usuarios, s_c = \arg_{s \in Itens} MaxR(c, s) \quad (2.3)$$

A principal preocupação de um sistema de recomendação é que a função de utilidade u geralmente não é definida no espaço todo de $C \times S$. A utilidade é representada tipicamente por avaliação e é inicialmente definida apenas nos itens previamente avaliados pelos usuários. Por exemplo, em um sistema de recomendação de filmes, os usuários inicialmente avaliam algum subconjunto de filmes que eles já assistiram. Considerando o espaço $C \times S$ como uma matriz de avaliações, muitas das avaliações não são fornecidos inicialmente, por isso a matriz é dita esparsa. Assim, o trabalho do sistema de recomendação é extrapolar u para todo o espaço $C \times S$. O caminho que é seguido para fazer estas extrapolações define a função de utilidade (Souza, 2011). Esta tarefa de extrapolação tem sido abordada de diversas maneiras (Shardanand e Maes, 1995; Balabanovic e Shoham, 1997; Resnick e Varian, 1997), como:

1. Especificando heurísticas que irão definir a função de utilidade empiricamente por meio da validação de seu desempenho.
2. Estimar a função de utilidade por meio da otimização de critérios de desempenho pré-definidos, como o erro médio quadrático (Adomavicius e Tuzhilin, 2005a).

Técnicas de Recomendação

Para implementar a função central, identificando os itens de interesse para o usuário, o Sistema de Recomendação deve prever qual item vale a pena recomendar. Para fazer isso, o sistema tem que ser capaz de prever a utilidade de alguns deles, ou comparar a utilidade de vários itens, e assim decidir quais itens recomendar baseados na comparação. O passo de predição pode não ser explícito no algoritmo de recomendação, mas pode ser aplicado ao modelo para descrever o rol geral de um Sistema de Recomendação (Ricci et al., 2011).

Para ilustrar o passo de predição, considera-se um algoritmo de recomendação simples, não-pessoalizado e que recomenda somente as canções mais populares. A lógica desse algoritmo de recomendação baseia-se na ausência de informações mais precisas sobre as preferências do usuário. É assim que

a recomendação será canções populares, ou seja, algo que é desejado (grande utilidade) por muitos usuários.

Burke (2007) apresentou uma taxonomia considerada uma forma clássica de diferenciar entre os seis diferentes tipos de abordagens para recomendação. Tais recomendações são:

Baseada em conteúdo: O sistema aprende a recomendar itens que são similares a outros que o usuário gostou no passado.

Filtrada Colaborativo: Esta abordagem recomenda ao usuário ativo os itens que outros usuários similares gostaram no passado. A similaridade entre os dois usuários é calculada baseada na similaridade do histórico dos usuários. O filtrado colaborativo é considerado como a técnica mais popular e mais implementada para sistemas de recomendação.

Demográfica: Este tipo de sistema recomenda itens baseado na estrutura demográfica do perfil do usuário. Assume-se que diferentes recomendações devem ser geradas por cada nicho geográfico. Por exemplo, os usuários são enviados a diferentes páginas Web baseados na língua do seu país.

Baseada em conhecimento: Recomenda itens baseados num domínio específico de conhecimento acerca de quantas características o usuário necessita, como preferências, e quão útil é o item para o usuário.

Baseada na comunidade: Recomenda sistemas baseados nas preferências de usuários amigos. Esta técnica segue o epigrama “Diga-me quem são seu amigos, e eu te direi quem você é”.

Sistemas de recomendação híbridos: Estão baseados numa combinação das técnicas mencionadas anteriormente. Estes tipos de sistemas combinam determinadas técnicas A e B e tentam usar as vantagens de A para consertar as desvantagens de B.

A continuação será descrita de forma mais detalhada a abordagem de filtragem baseada em conteúdo por ser a parte de maior interesse para a abordagem que se está propondo neste trabalho.

Filtragem Baseada em Conteúdo

Os sistemas baseados nesta técnica têm como objetivo gerar de forma automática, descrições dos conteúdos dos itens e comparar esta descrição de cada

item com a descrição dos interesses dos usuários ou com o histórico de consumo do mesmo, visando verificar se o item é ou não relevante para o usuário em questão (Balabanovic e Shoham, 1997).

A descrição de interesse do usuário é obtida por meio de informações fornecidas pelo próprio usuário, por meio de uma consulta, ou aprendendo com os itens que o usuário gostou anteriormente. Esta técnica é chamada de filtragem baseada em conteúdo por que o sistema realiza filtragem baseada na análise de conteúdo do item e no perfil do usuário (Herlocker, 2000; Ansari et al., 2000). As recomendações baseadas em filtragem de conteúdo são uma extensão das pesquisas relacionadas com filtragem de informação (Belkin e Croft, 1992). Em um sistema baseado em conteúdo, os objetos de interesse são definidos pelas suas características. Por exemplo, o *NewsWeeder* (Lang, 1995), um sistema de recomendação de notícias, usa as palavras dos textos para caracterizar os mesmos.

Muitas ferramentas que aplicam esta forma de filtragem utilizam a técnica de indexação de frequência de termos (Salton, 1989; Herlocker, 2000). Neste tipo de indexação, informações dos documentos e necessidades dos usuários são descritas por vetores com uma dimensão para cada palavra que ocorre na base de dados. Cada componente do vetor corresponde à frequência que uma respectiva palavra ocorre em um documento ou na consulta do usuário. Os vetores dos documentos que estão próximos aos vetores da consulta do usuário são considerados os mais relevantes para o usuário (Cazella, 2006). Outros exemplos de tecnologias que são aplicadas para filtragem baseada em conteúdo são índices de busca booleana, na qual a consulta constitui-se em um conjunto de palavras chaves unidas por operadores booleanos (Herlocker, 2000); sistemas de filtragem probabilística, na qual o raciocínio probabilístico é aplicado para determinar a probabilidade que um documento possui de atender as necessidades de informação de um usuário.

Os algoritmos de recomendação baseados em conteúdo aprendem os perfis de interesse dos usuários com base nas características presentes em cada um dos itens que esses usuários avaliaram previamente. Sendo assim, eles constroem os perfis dos usuários a partir do perfil dos itens. Em termos formais, sistemas de recomendação baseados em filtragem de conteúdo estimam a utilidade $u(c, s)$ de um item s para um usuário c baseado nas utilidades $u(c_i, s_i)$ indicadas pelo usuário c para os itens $s_i \in S$ que são similares ao item s (Adomavicius e Tuzhilin, 2005a).

2.4.2 Métricas de Avaliação

A qualidade de um sistema de recomendação pode ser avaliado comparando as recomendações de um conjunto de teste com escores marcados pelos usuários. Este tipo de sistemas são tipicamente avaliados por meio da utilização de três tipos de métricas de precisão: métricas de precisão na predição, na classificação e no ranqueamento.

Métricas de Precisão na Predição

Nas métricas de precisão preditivas (Herlocker et al., 2000), o conjunto de itens recuperados pelo sistema são diretamente comparados com aqueles que o usuário acha que o sistema deveria ter recuperado. A métrica mais comum na literatura é *Mean Absolute Error* (MAE). Esta métrica pode ser definida como a diferença média absoluta entre avaliações previstas e avaliações reais, e está representada na Equação 2.4

$$MAE = \frac{\sum_{u,i} |p_{u,i} - r_{u,i}|}{N} \quad (2.4)$$

Onde $p_{u,i}$ representa o valor previsto pelo sistema para o usuário u no item i ; $r_{u,i}$ é o escore atual, e N é o total de número de escores no teste.

Uma outra métrica muito utilizada é a *Root Mean Squared Error* (RMSE), que põe mais ênfase em erros absolutos maiores, e está definida como mostrada na Equação 2.5

$$RMSE = \sqrt{\frac{\sum_{u,i} (p_{u,i} - r_{u,i})^2}{N}} \quad (2.5)$$

As métricas de precisão preditivas tratam todos os itens da mesma forma. No entanto, para a maioria dos sistemas de recomendação, a principal preocupação é prever corretamente os itens que o usuário necessita.

Métricas de Precisão na Classificação

Nas métricas de precisão na classificação (Herlocker et al., 2000), o conjunto de itens recuperados pelo sistema são analisados em termos de quantidade de dados recuperados e a relevância que eles tem para o usuário. As métricas mais representativas na literatura são a *Precisão* (*Precision*) e a *CoBERTura* (*Recall*). A precisão pode ser definida como a quantidade de itens

relevantes recuperados de entre todo o conjunto de itens recuperados, e está representada na Equação 2.6

$$Precision = \frac{B_{rs}}{B_s} \quad (2.6)$$

Onde B_{rs} representa a quantidade de itens verdadeiramente relevantes que foram recuperados, e B_s é o total de itens recuperados. A cobertura pode ser definida como a quantidade de itens relevantes recuperados de entre todo o conjunto de itens relevantes existentes para esse domínio, e está representada na Equação 2.7

$$Cobertura = \frac{B_{rs}}{B_r} \quad (2.7)$$

Onde B_{rs} representa a quantidade de itens verdadeiramente relevantes que foram recuperados, e B_r é o total de itens considerados relevantes num determinado domínio.

Métricas de Precisão no Ranqueamento

Nas métricas de precisão no ranqueamento (Herlocker et al., 2000), o ranqueamento do conjunto de itens recuperados pelo sistema é comparado com o ranqueamento de itens estabelecido pelo usuário com o objetivo de determinar a correlação que existe entre eles. A métrica mais representativa na literatura é o coeficiente de correlação de Spearman representado na Equação 2.8

$$\rho = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)} \quad (2.8)$$

Onde $p_{u,i}$ representa o valor previsto pelo sistema para o usuário u no item i ; $r_{u,i}$ é o escore atual, e N é o total de número de escores no teste.

2.5 Considerações Finais

Neste capítulo foram explorados os conceitos e técnicas utilizadas para o desenvolvimento de este trabalho de Mestrado. Foram exploradas algumas técnicas importantes de Mineração de Textos tais como extração de similaridade entre documentos e agrupamento de dados. Apresentou-se também uma introdução à Lógica Fuzzy e à utilização de Ontologias e Ontologias Fuzzy como técnicas de representação estruturada de dados. A união dos conceitos e técnicas de Mineração de Textos, Lógica Fuzzy e Ontologias, permitiu a ela-

boração de uma abordagem para a utilização de Ontologias *fuzzy* de Domínio e Ontologias de Usuário no processo de recomendação baseado em conteúdo.

Trabalhos Correlatos

Atualmente, na literatura pode-se encontrar informação de sistemas que utilizam ontologias como base para obter seus resultados. Ressalta-se também a existência de várias ontologias crisp validadas, conhecidas e amplamente utilizadas. Por exemplo, ontologias de imagens, medicina, dispositivos móveis, filmes, redes de computadores, entre vários outros. Ao mesmo tempo, como explorado no Capítulo 2 existem algumas metodologias que permitem a criação de uma ontologia tradicional desde o início. No entanto, não existem muitos trabalhos que usem uma metodologia específica para poder transformar uma ontologia tradicional crisp em uma ontologia *fuzzy*.

No âmbito dos Sistemas de Recomendação, é comum utilizar os perfis de usuário para gerar recomendações mais personalizadas sobre o que o usuário realmente quer ou precisa. Alguns trabalhos utilizam como base ontologias para obter ou armazenar os dados de uma forma semanticamente mais rica. Entre as abordagens existem também várias formas de propagar o interesse do usuário á itens que podem ser de interesse para o usuário.

Neste capítulo são explorados os trabalhos relacionados com as metodologias que permitem construir uma ontologia *fuzzy*, principalmente as pesquisas que usam a similaridade entre conceitos da ontologia e são apresentados na seção 3.1. São exploradas também as ontologias de usuário como parte dos Sistemas de Recomendação e a influência que a propagação dos interesses tem nas recomendações, apresentado na seção 3.2.

3.1 Metodologias para Construção de Ontologias *Fuzzy*

As ontologias *fuzzy* estão baseadas na modificação de uma ontologia *crisp* existente. A modificação de uma ontologia tradicional em uma ontologia *fuzzy* é um processo inteiramente incremental, na qual são adicionados valores de pertinência entre as relações existentes da ontologia (Parry, 2004).

A continuação são detalhados em ordem cronológico os trabalhos mais ressaltantes na construção ou transformação de ontologias *crisp* em ontologias *fuzzy*.

3.1.1 Fuzzificação de uma ontologia tradicional CRISP (Parry, 2004)

No trabalho de Parry (2004) é proposta uma metodologia que permite fuzzificar uma ontologia tradicional *crisp*. Nesse trabalho, considera-se que o valor de pertinência pode ser associado principalmente de duas formas:

- Utilizando a preferência dos usuários;
- Usando um conjunto de documentos relacionados com o domínio da ontologia.

Fuzzificação baseado na preferência dos usuários

Nesta forma de fuzzificação, o autor considera que o valor de pertinência entre os conceitos de uma ontologia *fuzzy* podem ser usados para identificar a localização de termos similares dentro da ontologia. Cada usuário tem seus próprios valores de pertinência associados a cada termo dentro da ontologia. O processo de atribuir um grau de pertinência para cada conceito pode ser realizado durante o processo de busca de cada usuário ou de um grupo de usuários com características similares.

Fuzzificação baseado nos documentos

Nesta forma de fuzzificação, utilizou-se como base a ontologia hierárquica MeSH¹. Para o qual, os termos que ocorrem mais de uma vez na ontologia MeSH foram identificados em primeiro lugar. Após, foi verificado se um termo existe em várias localizações, de existir atribui-se um grau de pertinência proporcional para cada termo. Utilizou-se também um conjunto de documentos

¹<http://www.ncbi.nlm.nih.gov/mesh>

que contem em sua maioria os mesmos termos que foram considerados na ontologia MeSH. Após, é feita uma análise dos documentos e grau de pertinência de cada conceito é calculado usando algumas fórmulas baseadas principalmente na distancia entre as palavras.

3.1.2 FOGA - *Fuzzy Ontology Generation frAmework* (Quan et al., 2004)

No trabalho de [Quan et al. \(2004\)](#) é proposto um *framework* o qual é usado para a geração automática de ontologias *fuzzy* a partir de informação imprecisa. Esse *framework* está composto de três componentes:

1. *Fuzzy Formal Concept Analysis* - FFCA: o qual está encarregado da análise *fuzzy* de conceitos formais;
2. Módulo encarregado do agrupamento conceitual *fuzzy*;
3. Módulo encarregado da geração de relações hierárquicas.

No componente (1) FFCA, é construído um contexto formal *fuzzy* a partir de um banco de dados que apresenta informação imprecisa. Esse componente, é o encarregado de produzir conceitos formais *fuzzy* a partir do contexto formal *fuzzy*. O (2) agrupamento conceitual *fuzzy* agrupa os conceitos da malha de conceitos *fuzzy* produzida no FFCA e gera agrupamentos conceituais baseados na informação *fuzzy* contida na malha. Na (3) geração de relações hierárquicas são geradas as relações hierárquicas entre os agrupamentos conceituais com o intuito de construir uma hierarquia de conceitos.

Nesse trabalho os autores utilizam o *framework* proposto para gerar uma ontologia acadêmica para a web semântica de um determinado banco de dados composto de 1400 documentos científicos na área de Recuperação de Informação. Como resultado, a ontologia contem informação acadêmica distribuída numa hierarquia de acordo as áreas de investigação. Para a validação do *framework*, foi feita uma avaliação da qualidade da hierarquia de conceitos comparando os resultados obtidos pelo *framework* com os resultados produzidos por o COBWEB, que é considerada uma das melhores técnicas para o agrupamento conceitual. Na avaliação, mostra-se como o *framework* FOGA apresenta uma melhor performance na qualidade do agrupamento obtido.

3.1.3 Conhecimento *fuzzy* para integrar ontologias de domínio (Todorov et al., 2011)

Todorov et al. (2011) propuseram um algoritmo hierárquico que permite a representação de um conjunto de conceitos *fuzzy* utilizando uma ontologia de referência.

O processo de fuzificação proposto pelos autores consiste em atribuir escores para cada conceito base, calculando a similaridade entre o conceito base e os conceitos de referência. Os escores associados tomam valores entre 0 e 1 e são considerados como parte do conjunto *fuzzy*. Para obter a fuzificação, os autores propuseram o seguinte algoritmo hierárquico:

- Atribuem-se escores para cada vetor. Por exemplo, para cada nó folha atribui-se um grau de pertinência.
- Para cada nó não-folha que não contenha um documento por si mesmo é atribuído o valor máximo do escore do seus filhos.
- Se um nó não-folha tem instâncias associadas, o escore atribui-se considerando primeiro como base as instâncias e depois o máximo valor dos filhos.

Para fuzificar os conceitos base, os autores aplicaram o algoritmo hierárquico explicando anteriormente e usaram como referência a Wikipédia.

3.1.4 Roles de usuários na fuzificação de ontologias (Wallace et al., 2011)

No trabalho de Wallace et al. (2011) é apresentada a metodologia IKARUS-Onto que permite o desenvolvimento de ontologias *fuzzy*. Esta metodologia assume a existência de uma ontologia tradicional e descreve as ações que devem ser levadas em consideração para gerar uma versão *fuzzy* da ontologia. Essa metodologia está baseada em quatro fases:

- Adquisição da ontologia crisp: que utiliza o conhecimento contido na ontologia tradicional e estabelece as bases para desenvolver a fuzificação.
- Estabelecimento das necessidades *fuzzy*: que Verifica em que lugar da ontologia trabalhara-se com a imprecisão e verifica-se também se o domínio atual inclui imprecisão.

- Definição dos elementos da ontologia *fuzzy*: Nesta etapa identificam-se as áreas da ontologia onde a imprecisão existe e calculam-se os graus de pertinência para os termos da ontologia que representam imprecisão. Esta etapa não pode ser realizada sem a supervisão de um engenheiro da ontologia ou um experto de domínio.
- Formalização dos elementos *fuzzy*: refere-se à seleção da linguagem na qual será expressada a ontologia.
- Validação da ontologia *fuzzy*: Verifica a consistência, exatidão e precisão da ontologia gerada. Essa validação é feita com ajuda do engenheiro de ontologia e o experto do domínio.

Finalmente, ressalta-se que para a fuzzificação de uma ontologia baixo esta abordagem é preciso a ajuda em cada etapa de um engenheiro da ontologia e um experto do domínio.

3.2 Sistemas de Recomendação baseados em Ontologia de Usuário e Propagação do Interesse

As ontologias provaram ter um melhor desempenho que outras abordagens quando tenta-se processar a representação da informação. As ontologias podem ter uma forma simples como uma taxonomia ou podem ter associados vocabulários com terminologia baseada em definições da linguagem natural. As ontologias baseadas nos perfis de usuários estão sendo bastante aplicadas na representação do contexto dos usuários e em personalização das aplicações (Satler et al., 2010).

A ideia geral dos Sistemas de Recomendação baseados em conteúdo é observar as buscas feitas pelos usuários e recomendar páginas relacionadas com as buscas efetuadas. A continuação são detalhados em ordem cronológico os trabalhos mais ressaltantes que utilizam as ontologias de usuário ou os perfis de usuário e a propagação do interesse.

3.2.1 Aprendendo e inferindo uma ontologia de usuário para serviços semânticos web personalizados (Jiang e Tan, 2006)

No trabalho de Jiang e Tan (2006), é proposta a criação de uma ontologia baseada no usuário. Essa ontologia está formada por conceitos e relações semânticas que permitem representar o interesse do usuário. Para poder

aprender e armazenar a mudança dos interesses do usuário, foi considerado o procedimento de propagação de ativação (*spreading activation*).

Os conceitos de interesse do usuário são armazenados junto com a frequência. Essa frequência está estabelecida pela quantidade de vezes que o usuário procurou o termo na Web. Para determinar as relações de interesse, calcula-se um valor para cada relação armazenada no histórico de busca. Para poder calcular esse valor, os autores utilizaram uma solução Bayesiana.

No sistema proposto pelos autores, o usuário ingressa à consulta de um termo. Após, o mecanismo de busca devolve uma lista inicial de documentos recuperados utilizando os métodos clássicos baseados em palavra-chave. Essa lista de documentos tem conceitos associados a cada documento e uma frequência de repetição para cada documento. Os conceitos junto com suas frequências colocam-se na forma de um vetor. Quando o sistema recebe o vetor é realizado o procedimento de propagação de ativação. Essa propagação utiliza a ontologia de usuário para inferir os conceitos de relevância. Após a propagação, é gerado um outro vetor que contém os valores dos conceitos propagados. Após, esses valores são combinados com o fator de interesse obtido para cada conceito. Com esses valores gera-se um *ranking* de conceitos. Esse *ranking* considera o intervalo de tempo desde que a última consulta foi realizada pelo usuário. Esse fator de tempo é considerado como a função de decaimento que visa prevenir a saturação do interesse de um termo dentro da ontologia. Finalmente, os documentos com um escore alto são colocados no topo da lista para apresentar esses elementos ao usuário.

3.2.2 Personalização dos mecanismos de busca com perfis ontológicos de usuários (Sieg et al., 2007)

Sieg et al. (2007) apresentam uma abordagem para personalizar os sistemas de busca. Isso envolve a construção de modelos do contexto do usuário como perfis ontológicos que tem associado um interesse para os conceitos da ontologia de domínio. Os autores usaram também a teoria de propagação da ativação para manter o interesse do usuário baseados no seu comportamento.

O objetivo principal dos autores é utilizar o contexto do usuário para personalizar os resultados da busca. Cada ontologia de usuário é inicialmente uma instanciação da ontologia de domínio. Cada conceito na ontologia do usuário é marcada com um escore de interesse que tem um valor inicial de 1. Enquanto o usuário interatua com o sistema selecionando documentos ou fazendo buscas, a ontologia do usuário é atualizada e as marcações para cada

conceito são modificadas utilizando a propagação de ativação.

A propagação da ativação é utilizada para atualizar de forma incremental o escore de interesse dos conceitos na ontologia do usuário. Considerando que os conceitos estão organizados de forma hierárquica e derivados da ontologia de domínio, os pesos de interesse são calculados para as relações que existem entre os conceitos e sub-conceitos. Para cada iteração do algoritmo, o valor de cada conceito na ontologia de usuário vira zero. Obtém-se um vetor de documentos e compara-se com o vetor de conceitos utilizando a similaridade do cosseno. Aqueles conceitos que tem uma similaridade maior que zero são adicionados a uma lista. Após, é calculado o valor de ativação para cada elemento da lista e os conceitos com os maiores valores são removidos da lista e processados. Para esses valores extraídos e propagado o interesse para os nós relacionados tais como nós irmãos ou nós pais.

Os autores demonstraram que ao incluir a propagação do interesse dentro de uma ontologia de usuário e ao incluir uma ontologia baseada nas preferências dos usuários dentro de sistemas de busca estes melhoram em termos de número de itens de interesse que são apresentados para o usuário depois de realizar uma busca.

3.2.3 Perfil de usuário ontológico em recomendações personalizadas em *e-Commerce* (He e Fang, 2008)

He e Fang (2008) explorou o perfil do usuário usando ontologias para melhorar as recomendações em *e-commerce*. O perfil do usuário é representado em classes ontológicas que consideram o interesse do usuário associado a cada conceito. Nessa ontologia, foram considerados enlaces do tipo “é-um” entre os conceitos. Os interesses do usuário são classificados como termos compridos e termos curtos de interesse e cada interesse é armazenado usando uma estrutura de árvore. Depois que um usuário loga no sistema, a árvore de interesses associada a esse usuário é obtida. Essa árvore pode ser observada pelo usuário para que ele possa manualmente personalizar a lista de preferências relacionadas com os termos curtos. Os termos compridos da árvore são obtidos das ações que realiza o usuário com o sistema dentro de um período de tempo.

Após a construção do perfil do usuário, utiliza-se a inferência da ontologia para encontrar interesses ocultos para o usuário. Esses interesses ocultos são conceitos no qual usuário poderia estar interessado mas não de uma forma óbvia. Para encontrar os interesses ocultos o sistema calcula a distância entre

dois pares de conceitos dentro da ontologia.

3.2.4 Busca de informação contextual baseada no perfil ontológico de usuário (Mohammed et al., 2010)

No trabalho de Mohammed et al. (2010), é construído um perfil ontológico de usuário que descreve as preferências dos usuários baseados nos contextos. Para isso, os autores utilizaram uma análise semântica dos *logs* dos usuários para aprender os interesses considerando o tempo.

Nessa abordagem, o perfil do usuário é considerada como uma instância da ontologia de domínio, na qual cada conceito representa o interesse do usuário. A ontologia do perfil do usuário é gerada fazendo um mapeamento entre a ontologia de domínio e o contexto do usuário. Para a construção da ontologia de domínio foram utilizados um conjunto de documentos associados a cada conceito. Todos os documentos que pertencem a um documento foram unidos para criar uma coleção de documentos por cada contexto, produzindo desta forma um vetor de características relacionadas para cada conceito dentro da ontologia.

Para obter os interesses dos usuário ao longo de certo tempo, os arquivos “.log” foram utilizados. Cada página armazenada nesse arquivo foi explorada e foi extraído o conteúdo associado, provendo desta forma uma lista dos interesses do usuário relacionados com as buscas feitas no passado. Esses termos de interesses foram mapeados na ontologia de domínio. Após, foi calculada a similaridade entre os conceitos mapeados da ontologia de usuário e a ontologia de domínio para obter uma lista dos conceitos mais representativos para o usuário.

Após, a ontologia do perfil do usuário foi utilizada num sistema de busca, no qual o ranking da lista de termos de interesse do usuário foram consideradas para ordenar os resultados de uma determinada busca. Os documentos recuperados de uma busca foram filtrados e comparados com a lista de interesse do usuário utilizando a medida de similaridade do cosseno.

Os experimentos mostraram que a abordagem proposta pelos autores melhora os resultados de busca já que utiliza os interesses do usuário previamente calculados para ordenar os documentos extraídos.

3.2.5 Abordagem de ontologia *fuzzy* para representar perfis de usuário em ambientes de *e-learning* (Satler et al., 2010)

No trabalho de Satler et al. (2010), é explorado como uma ontologia *fuzzy* pode representar os perfis de usuário em ambientes de *e-learning*. Segundo os autores, a representação ontológica do usuário melhora o desempenho em tarefas como filtragem e Recuperação de Informação.

A abordagem proposta pelos autores inclui várias etapas que permitem o processamento dos dados para gerar automaticamente uma ontologia *fuzzy* baseada no perfil do usuário utilizando uma coleção de documentos previamente selecionada. O processo de construção da ontologia foi dividida em:

- Pré-processamento linguístico: O objetivo é extrair a informação textual em forma de palavras individuais dos objetos relacionados com cada usuário.
- Indexação de termos (chamado de pré-ontologia): o objetivo é prover uma estrutura de índices que contenha informação acerca de todos os termos gerados na etapa anterior. Depois de fazer uma análise dos documentos, foram extraídas as características “TermID” e uma lista das características de cada documentos. Essa lista de características contem “DocumentID” e “Número de ocorrências do termo mais frequente no documento”.
- Extração dos termos relevantes para o usuário: Observou-se três níveis de relevância: termos Relevantes, termos sub-relevantes e outros termos. Para obter esses valores de relevância os autores utilizaram as medidas de tf-idf, tf e idf. Os termos que tenham maior relevância foram considerados como “Termos relevantes”, os termos com relevância maiores à média foram considerados como “Termos sub-relevantes” e os termos com graus menores que a média foram considerados como “Outros termos”.
- Geração da ontologia de usuário: A ontologia foi considerada como um conjunto de grafos direcionados no qual cada nó representa um item e as arestas denotam que os termos estão relacionados. Para calcular o valor de relacionamento *fuzzy* os autores usaram o algoritmo proposto por Widyantoro e Yen (2001).
- Atualização do perfil: Para cada termo da ontologia é calculado um peso de relevância e são atualizados considerando: (1) se o termo não existe

nos termos relevantes associados ao usuário então o valor não muda; e (2) se o termo existe nos termos relevantes do usuário então o termo é atualizado com o novo valor calculado.

3.3 Considerações Finais

Neste capítulo foram apresentados trabalhos que detalham metodologias para permitir a construção de ontologias. Foram explorados também diversos trabalhos que utilizam os perfis do usuário em forma de ontologia e a propagação dos interesses para poder obter resultados mais específicos para cada usuário.

Esses trabalhos foram utilizados como orientação para a proposta da abordagem de recomendação baseada em ontologia *fuzzy* de domínio e ontologia de preferências de usuário apresentada no seguinte Capítulo.

ObRA: Uma Abordagem para Recomendação baseada em Ontologias de Domínio e em Preferências de Usuário

Este capítulo apresenta a abordagem ObRA (Ontologies-based Recommendation Approach), a qual define estratégias para a construção de Sistemas de Recomendação baseados em ontologia *fuzzy* de domínio e em ontologia crisp de preferência de usuário. Esta abordagem baseia-se nos conceitos e nas técnicas descritas no Capítulo 2. O objetivo da proposta é desenvolver uma abordagem que permita determinar quais elementos sugerir a um usuário baseando-se na similaridade de elementos dentro de um determinado domínio. Para atingir esse objetivo, na abordagem proposta são combinadas as concepções de ontologias *fuzzy* de domínio baseadas em similaridade e em ontologias de preferências de usuário. Este capítulo está organizado da seguinte forma: A Seção 4.1 fornece uma visão geral da abordagem de recomendação baseada em ontologias *fuzzy* de domínio e em ontologia crisp de preferência do usuário, além de seus principais elementos e etapas, cujas atividades são detalhadas nas seções subsequentes 4.1.1 e 4.1.2. A Seção 4.2 apresenta um estudo de caso da aplicação da abordagem no domínio do comércio de dispositivos móveis.

4.1 Visão Geral da Abordagem ObRA

Considerando as ideias apresentadas na revisão bibliográfica, este trabalho investiga e propõe uma abordagem de recomendação que tem por objetivo fornecer sugestões mais precisas para os usuários com base nas suas preferências. A abordagem proposta é baseada em conteúdo, na qual a base da recomendação é o relacionamento entre os documentos que representam cada conceito. Esse relacionamento foi feito utilizando ontologias *fuzzy* do domínio baseadas na similaridade das instâncias.

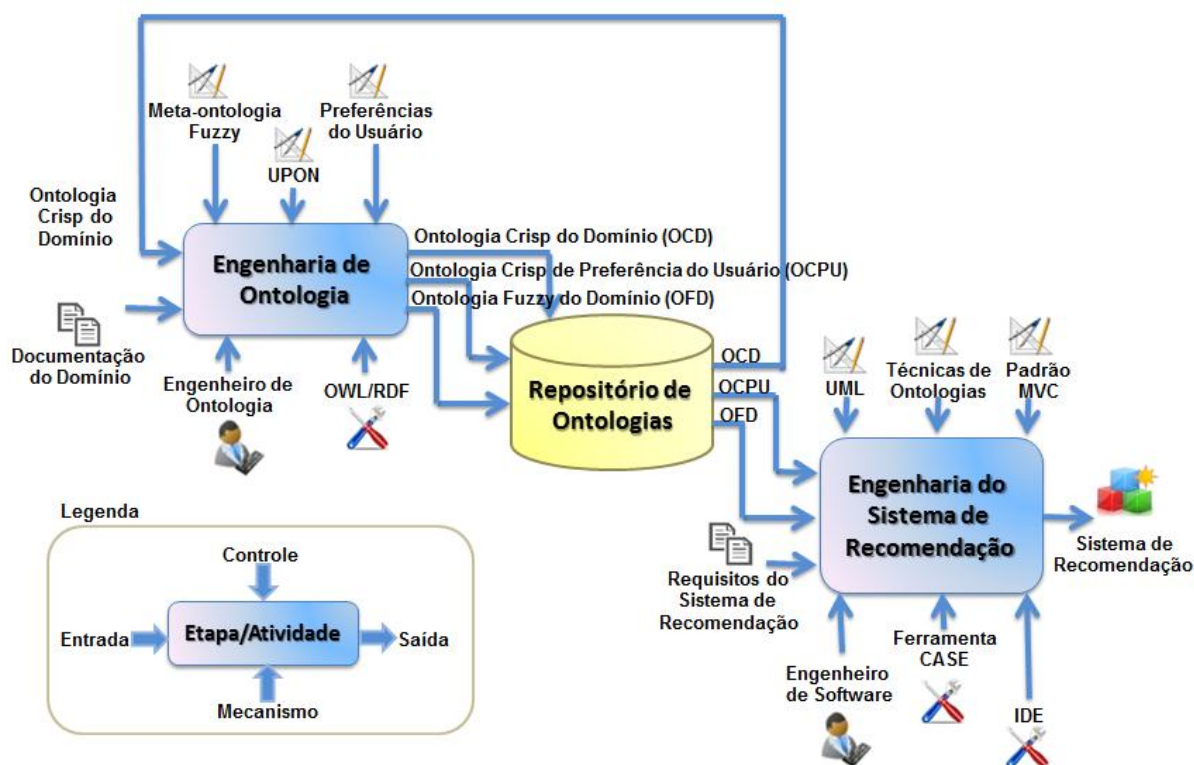


Figura 4.1: Visão Geral em alto nível da Abordagem de Recomendação baseada em Ontologias *Fuzzy*

Na Figura 4.1, é ilustrado o diagrama *Structured Analysis and Design Technique* (Ross, 1977) que representa a visão geral da abordagem. Como se pode observar, a abordagem de recomendação baseada em ontologias é realizada em duas etapas: Engenharia de Ontologia e Engenharia do Sistema de Recomendação. Essa divisão em duas etapas baseia-se nas abordagens da Engenharia de Software conhecidas como linhas de produção de software (*Software Product Lines - SPL*) (Linden et al., 2007). Nesse diagrama, os retângulos representam as atividades da abordagem, as setas que entram pelo lado

esquerdo dos retângulos representam as entradas de dados e as que saem pelo lado direito representam as saídas geradas em cada atividade. As setas do lado superior representam os controles que orientam a execução de cada atividade, já as setas do lado inferior representam os participantes, as ferramentas e os mecanismos que executam ou automatizam a execução das atividades.

A Engenharia de Ontologia é responsável por desenvolver os artefatos que apoiam o desenvolvimento da Engenharia do Sistema de Recomendação. Esses artefatos compreendem uma ontologia *fuzzy* de domínio específico que permita determinar os relacionamentos de similaridade entre os conceitos, além de uma ontologia que permita administrar as preferências de cada usuário. Os artefatos provenientes da etapa de Engenharia de Ontologia são armazenados no Repositório de Ontologias para serem posteriormente utilizados. Na Engenharia do Sistema de Recomendação, o Engenheiro de Software parte dos requisitos da aplicação e, o com apoio das ontologias produzidas na Engenharia de Ontologia, desenvolve as aplicações de recomendação. Nas seções seguintes os processos de Engenharia de Ontologia e de Engenharia do Sistema de Recomendação serão explorados detalhadamente.

4.1.1 Engenharia de Ontologia

A engenharia de ontologia incorpora decisões sobre como representar uma ampla seleção de objetos e relações dentro de uma ordem lógica, levando a um modelo de nível ontológico. No diagrama SADT da Figura 4.2 são ilustradas as etapas da engenharia de ontologia: construção de ontologia crisp do domínio, construção de ontologia crisp de preferências do usuário e fuzzificação de ontologias crisp do domínio.

Construção de Ontologia Crisp do Domínio

A primeira etapa, Construção de Ontologia Crisp do Domínio é realizada quando não é fornecida uma Ontologia Crisp do Domínio como parâmetro de entrada e utiliza a documentação do domínio como parâmetro de entrada. Nessa etapa, o engenheiro de ontologia, apoiado nas linguagens para representação de ontologias owl/rdf, e seguindo a metodologia UPON (De Nicola et al., 2009), gera como resultado uma Ontologia Crisp do Domínio que pode ser utilizada como parâmetro de entrada para as próximas etapas.

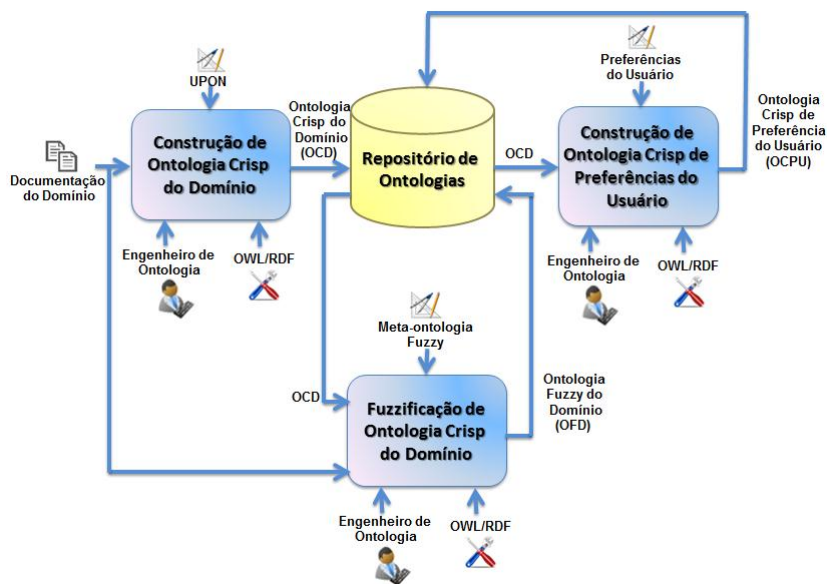


Figura 4.2: Visão Geral da Engenharia de Ontologias

Construção de Ontologia Crisp de Preferências do Usuários

A segunda etapa no processo da Engenharia de Ontologia compreende a modelagem de uma ontologia crisp de preferências de usuário, que neste trabalho baseia-se na abordagem de [Jiang e Tan \(2006\)](#), na qual essa ontologia crisp de preferências de usuário é tratada como uma instanciação da ontologia de domínio que armazena os valores de interesse associados a cada conceito para cada usuário. Esses valores de interesse são obtidos por meio da interação do usuário com os dados presentes na ontologia, a qual é dada por meio de um módulo de recomendação de elementos associados à ontologia. Durante cada interação é possível determinar o conceito com o qual o usuário está interagindo e, assim, atribuir um valor de interesse para esse conceito. O cálculo do valor de interesse para um conceito é dado pela Equação 4.1.

$$I(N) = I_o(N) + I_s(N) \quad (4.1)$$

A Equação 4.1 calcula o interesse do usuário I no conceito N , considerando que o conceito N recebe um *feedback* direto do usuário. O *feedback* é obtido por meio da interação do usuário com o módulo de recomendação, o qual influencia diretamente o valor de interesse (peso) de um conceito. A função $I_o(N)$ representa o valor anterior do interesse do conceito N . A função $I_s(N)$ representa o valor atual do interesse do conceito N e é calculado como ilustrado na Equação 4.2.

$$I_s(N) = \frac{l(N) + 1}{max + 1} * f(N)^b \quad (4.2)$$

A fim de calcular o valor atual de interesse do conceito N , a Equação 4.2 considera uma constante b com um valor de 0.5. A função $l(N)$ representa o nível do conceito N dentro da ontologia, max representa a altura máxima da ontologia e a função $f(N)$ representa o valor do *feedback* do usuário. Uma vez calculado o valor de interesse do conceito selecionado pelo usuário, esse deve ser propagado para os conceitos diretamente relacionados baseado na estrutura da ontologia. Para realizar a propagação desses valores, utiliza-se a abordagem de propagação de interesses adotada por [Cena et al. \(2011\)](#), na qual o interesse propagado é definido pela Equação 4.3.

$$I(M) = I_o(M) + I_p(M) \quad (4.3)$$

Nessa Equação, o valor de interesse é propagado para os conceitos M . A função $I_o(M)$ representa o valor do interesse anterior do conceito M e, inicialmente, o valor é 0. A função $I_p(M)$ representa o valor do interesse propagado do conceito M e é calculado como ilustrado na Equação 4.4.

$$I_p(M) = \frac{e^{-k*d(N,M)}}{1 + \log(1 + n(M))} * I_s(N) \quad (4.4)$$

Para calcular o valor de interesse propagado no conceito M , a Equação 4.4 considera um coeficiente k com um valor de 0.8. O valor $d(N, M)$ é a distância que existe entre o conceito inicial N e o conceito propagado M . A função $n(M)$ representa o número de vezes que o valor de interesse do conceito M foi modificado devido a uma ação direta do usuário.

Considerando que a propagação de interesses é feita por meio dos relacionamentos de similaridade existentes na ontologia, a distância entre os conceitos N e M é definida como o grau de dissimilaridade (1-similaridade) existente entre os conceitos, devido ao fato do conceito de distância ser considerado inversamente proporcional ao conceito de similaridade.

Uma vez terminada a etapa de Engenharia de Ontologia, os artefatos resultantes, no caso, a ontologia *fuzzy* de domínio e a ontologia *crisp* de preferências de usuário, são armazenados no repositório para uso posterior na etapa da construção do Sistema de Recomendação.

Fuzzificação de uma Ontologia

O processo de construção de ontologias apresenta uma complexidade estrutural e lógica comparável à produção de um artefato de software, pois o perfil do engenheiro de ontologia consiste da capacidade dos engenheiros do conhecimento e dos especialistas do domínio.

Dentre as abordagens para a construção de ontologias, a metodologia *Unified Process for ONtology* (UPON) (De Nicola et al., 2009) foi a escolhida como guia de desenvolvimento deste trabalho, pois essa utiliza as vantagens do *Unified Process* (UP) e do *Unified Modeling Language* (UML), que permite o desenvolvimento de uma ontologia de domínio eficiente.

A metodologia UPON foi desenvolvida com o objetivo de estabelecer uma série de etapas estruturadas para a modelagem e a criação de uma ontologia desde o momento da sua concepção (De Nicola et al., 2009).

No âmbito do trabalho com ontologias *fuzzy*, essa metodologia foi adaptada, por meio da adição de novas etapas que permitam inserir conceitos e relacionamentos *fuzzy* aos fluxos de requisito, análise e projeto, como ressaltado na Figura 4.3. Essas etapas são definidas com o objetivo de permitir a criação de uma ontologia *fuzzy* ou de realizar a fuzzificação de uma ontologia crisp existente. No restante deste trabalho, essa adaptação recebe o nome de *Unified Process for Fuzzy Ontology Building*, e visa auxiliar o engenheiro de ontologia no processo de fuzzificação de ontologias.

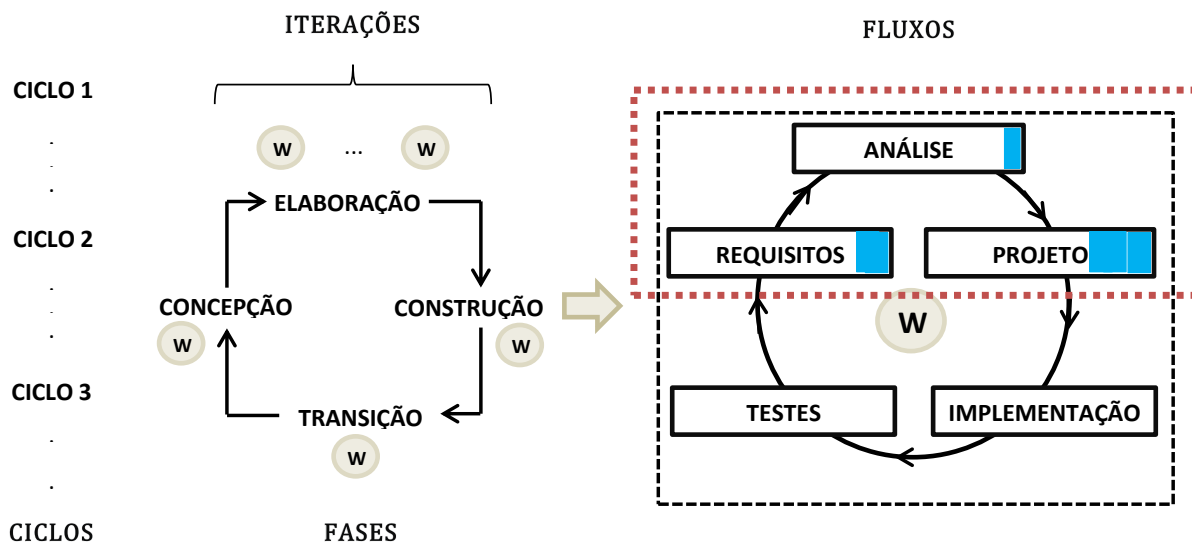


Figura 4.3: Representação da metodologia UPON de De Nicola et al. (2009), ressaltando os fluxos adaptados para permitir a criação ou fuzzificação de uma ontologia

Na Figura 4.3, são ilustrados os ciclos, as fases e os fluxos da metodologia UPON e a adaptação UPFON é ressaltada nos fluxos de requisitos, de análise e de projeto. No fluxo de requisitos, as necessidades semânticas são especificadas e o conhecimento é codificado na ontologia. No de análise, todos os requisitos da ontologia identificados no fluxo de requisitos são refinados e estruturados. No de projeto, os conceitos identificados no fluxo de análise são refinados e as relações entre eles são identificadas e além disso os relacionamentos *fuzzy* de similaridade são estabelecidos e seus valores são expressos na ontologia.

A seguir, encontra-se uma descrição de cada uma das etapas inseridas como parte da adaptação UPFON. As etapas correspondentes à metodologia UPON foram descritas na Seção 2.3.2.

1. Fluxo de Requisitos: O objetivo principal desse fluxo é especificar os requisitos semânticos necessários para a construção da ontologia e estabelecer a visão do conhecimento do usuário para, assim, codificar tal conhecimento na ontologia. De acordo com Jacobson et al. (1999), essa atividade requer a cooperação entre projetistas, engenheiros do conhecimento e usuários, a fim de identificar os objetivos da ontologia do ponto de vista do usuário. O fluxo de requisitos recebe como parâmetros de entrada as entrevistas com os especialistas do domínio e a documentação específica sobre o domínio da aplicação. Nesse fluxo são adicionadas duas atividades, como ilustrado na Figura 4.4.
 - Determinar o propósito da fuzzificação, ajuda a obter uma melhor compreensão do domínio de interesse, analisando a necessidade de utilizar uma ontologia *fuzzy* no domínio específico;
 - Determinar os contextos envolvidos na ontologia, permite estabelecer sobre quais contextos acontecerá o relacionamento entre os conceitos da ontologia. Por exemplo, em uma ontologia de dispositivos móveis, os contextos poderiam estar determinados pelas características dos dispositivos ou pelo preço dos dispositivos.

O fluxo de requisitos tem como dados de saída: o léxico de aplicação, as questões de competência e os modelos de caso de uso.

2. Fluxo de Análise: O objetivo principal desse fluxo é fornecer um conjunto de atividades para a estruturação e o refinamento dos requisitos da ontologia identificados no fluxo de requisitos. Esse fluxo recebe como

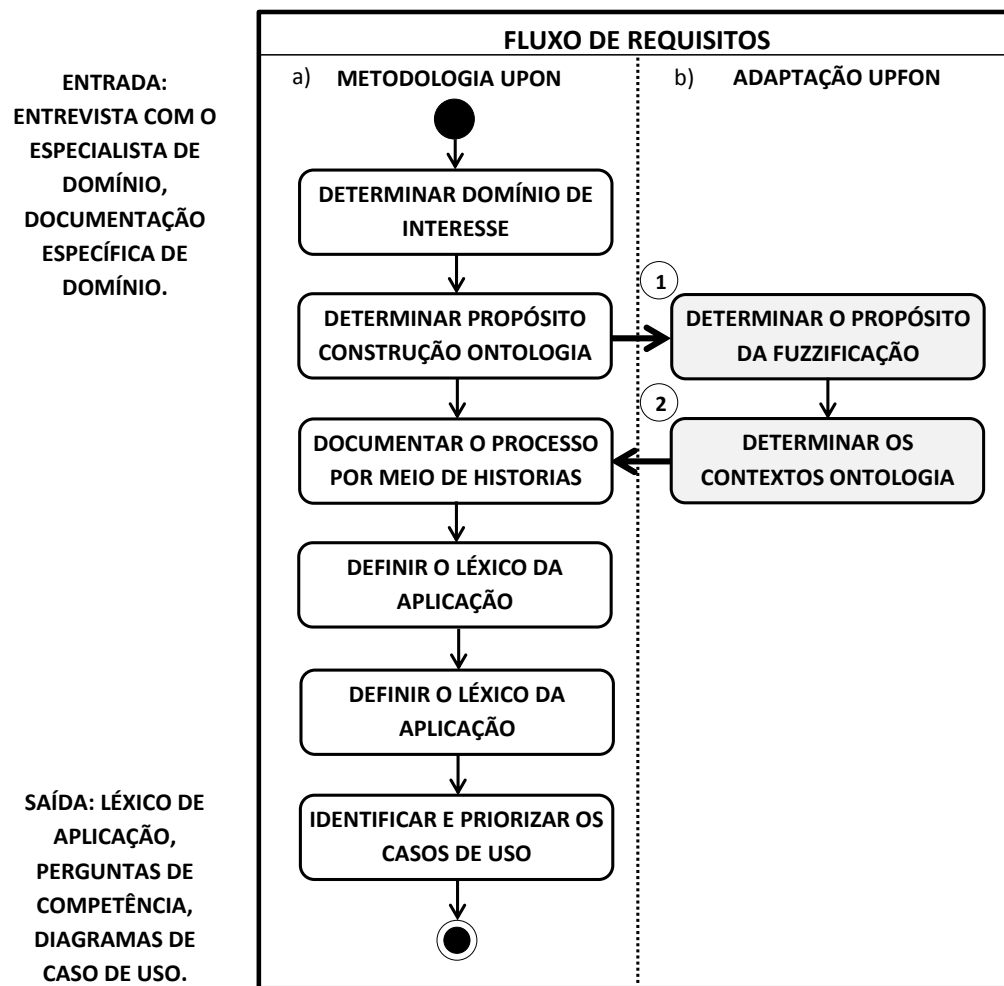


Figura 4.4: Atividades do Fluxo de Requisitos. a) Atividades da Metodologia UPON. b) Atividades da Adaptação UPFON

parâmetros de entrada um conjunto de documentos relacionados à área do domínio e ao léxico da aplicação, obtido como parâmetro de saída no fluxo de requisitos. Com esses parâmetros de entrada, o léxico da aplicação é enriquecido por meio de um domínio mais geral, a fim de construir o léxico de referência. Esse último é utilizado para produzir o glossário de referências, no qual são adicionadas definições para os termos. Nesse fluxo é adicionada uma atividade, conforme ilustrado na Figura 4.5.

- Adquirir recursos específicos do domínio, na qual tem como objetivo coletar documentos do domínio que forneçam informações relacionadas aos contextos definidos no fluxo de requisitos.

O fluxo de análise tem como dados de saída o léxico de domínio, o léxico

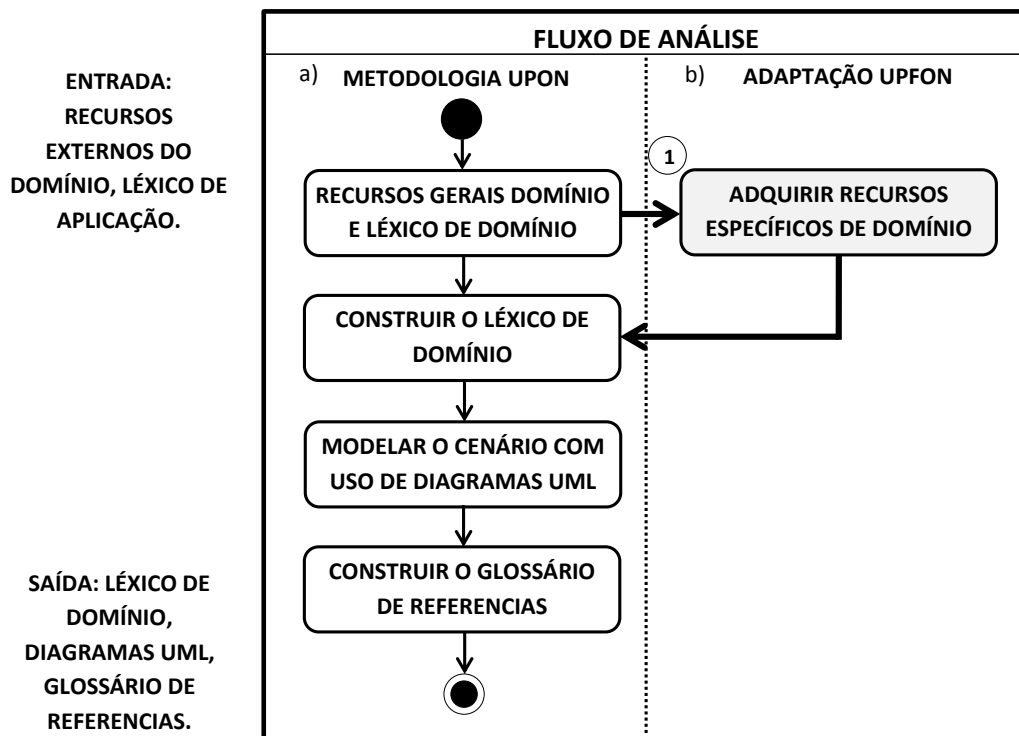


Figura 4.5: Atividades do Fluxo de Análise. a) Atividades da Metodologia UPON. b) Atividades da Adaptação UPFON

de referências, os diagramas UML e o glossário de referências.

3. Fluxo de Projeto: O principal objetivo do fluxo do projeto é definir uma estrutura ontológica baseada no conjunto de dados presente no glossário de referências. A estruturação dos dados é feita por meio da organização hierárquica dos termos e da definição de atributos e axiomas para cada um deles. Nesse fluxo, realiza-se o refinamento das entidades e dos processos identificados durante o fluxo de análise, assim como a identificação dos relacionamentos semânticos existentes entre cada uma dessas entidades. O relacionamento semântico entre dois conceitos é dado pela existência de um grau de similaridade entre tais conceitos em um contexto específico, por exemplo, a similaridade no contexto de características ou no preço de um item. Nesse fluxo são adicionadas quatro atividades, conforme ilustrado na Figura 4.6.

- Pré-Processamento dos Documentos Específicos do Domínio: O relacionamento semântico entre os conceitos é determinado pelo grau de similaridade que existe entre os documentos a serem avaliados. Na metodologia proposta, esses documentos devem ter uma estrutura

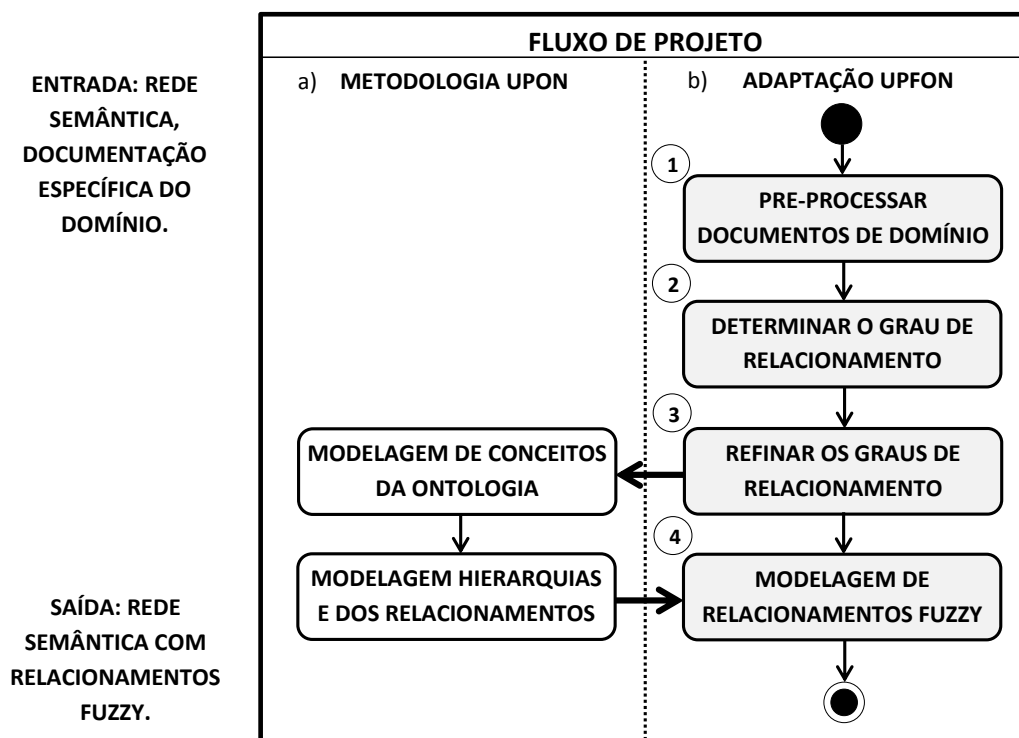


Figura 4.6: Atividades do Fluxo de Projeto. a) Atividades da Metodologia UPON. b) Atividades da Adaptação UPFON

pré-definida do tipo atributo-valor. Por exemplo, no contexto de similaridade por características no domínio de veículos, alguns dados do documento poderiam ser: “tipo:tipo_veículo”, “modelo:nome_modelo”, “cor:cor_veículo”.

A obtenção desse documento ocorre por meio da depuração de dados dos documentos específicos do domínio. Para a realização dessa depuração existem diversas técnicas de mineração de texto (e.g. *stopwords*¹, n-gramas, entre outras) que permitem a adaptação do conteúdo do documento ao resultado que se procura. O documento resultante dessa depuração influencia diretamente no processo de cálculo do grau de similaridade dos conceitos, pelo qual é importante verificar que os documentos específicos do domínio, coletados no fluxo de requisitos, permitam a identificação desses dados atributo-valor. Também é importante ressaltar que os valores obtidos nessa etapa devem ser validados por um especialista do domínio.

¹palavras que se repetem inúmeras vezes no decorrer do texto ou palavras sem relevância aparente para o entendimento do texto

Logo, a atividade de pré-processamento dos documentos específicos do domínio tem como objetivo principal extrair uma representação vetorial das ocorrências das palavras em cada documento. Uma forma de obter essa representação é a partir do modelo de dados conhecido como *bag-of-words*, que consiste em uma tabela de tipo atributo-valor na qual as linhas correspondem aos documentos e as colunas aos atributos ou palavras pertencentes ao conjunto de documentos. Cada célula tem um valor que relaciona a frequência do atributo ao documento. Dessa forma, um documento pode ser representado em um espaço vetorial.

- **Determinação do Grau de Relacionamento dos Conceitos:** O objetivo dessa atividade é determinar o valor do grau de relacionamento semântico existente entre os conceitos. Na metodologia proposta, o tipo de relacionamento semântico utilizado é determinado pela similaridade existente entre dois conceitos. O valor do grau de relacionamento de similaridade entre os conceitos pode ser determinado de acordo com os contextos identificados no fluxo de requisitos ou pela estrutura hierárquica da ontologia. Nos casos em que os graus de relacionamento dependem dos contextos identificados é estabelecido um grau de relacionamento diferente para cada um desses contextos.

O relacionamento semântico de similaridade determinado pela estrutura hierárquica da ontologia segue as etapas adotadas em (Su et al., 2010). Nessa abordagem, o grau de relacionamento é calculado por meio da estrutura da ontologia em si, pela junção dos valores de similaridade baseados na distância, na profundidade e na densidade entre dois conceitos da ontologia.

Para o cálculo do valor de similaridade baseado na distância utiliza-se a Equação 4.5, na qual $distance(a, b)$ representa a distância existente entre os conceitos a e b na ontologia, e α representa o valor de relevância determinado para esse fator. O valor da distancia entre a e b é determinado considerando a quantidade de conceitos que separam o conceito a do b . O valor de relevância deve ser determinado pelo especialista de domínio, já que esse indica a importância que o fator representa no resultado final.

$$sim_{distance}(a, b) = \frac{\alpha}{distance(a, b) + \alpha} \quad (4.5)$$

Para o cálculo do valor de similaridade baseado na profundidade é utilizada a Equação 4.6, na qual $depth(a)$ representa a profundidade do conceito a em relação à estrutura da ontologia. O valor da profundidade é determinado pela distância existente entre o conceito a e o conceito $raiz$ (nó raiz).

$$sim_{depth}(a, b) = \frac{depth(a) + depth(b)}{|depth(a) - depth(b)| + 1} \quad (4.6)$$

Para o cálculo do valor de similaridade baseado na densidade utiliza-se a Equação 4.7, na qual $density(a, b)$ representa a densidade de nós existentes entre os conceitos a e b da ontologia. O valor da densidade entre a e b é determinado pela Equação 4.8, onde n representa o número de conceitos filhos do conceito imediato superior em comum e m representa a profundidade do mesmo.

$$sim_{density}(a, b) = \frac{1}{density(a, b)} \quad (4.7)$$

$$density(a, b) = \frac{n}{m} \quad (4.8)$$

Após o cálculo dos valores de similaridade por distância, profundidade e densidade, é possível determinar o valor final de similaridade entre os conceitos a e b da ontologia. O valor final de similaridade é determinado pela Equação 4.9, na qual cada valor de similaridade calculado é multiplicado por um valor de relevância (θ , β , γ , onde $\theta + \beta + \gamma = 1$). O valor de relevância para cada variável indica a importância que cada uma delas representa no resultado final da similaridade.

$$sim_1(a, b) = \theta * sim_{distance}(a, b) + \beta * sim_{depth}(a, b) + \gamma * sim_{density}(a, b) \quad (4.9)$$

Outra forma de obter o valor do grau de similaridade do relacionamento é por meio do relacionamento semântico de similaridade determinado pelos contextos. Esse tipo de relacionamento baseia-se na similaridade existente entre os documentos específicos de domínio coletados no fluxo de requisitos. Vale ressaltar que tais documentos relacionam-se ao contexto sobre o qual se quer obter o relacionamento.

O valor de similaridade entre documentos é determinado por meio da comparação das representações vetoriais de cada um deles. Na metodologia proposta, essas representações são obtidas como resultado da atividade de pré-processamento de documentos. Um método bem estabelecido na área de Recuperação de Informação para calcular a similaridade entre representações vetoriais de documentos é a medida da similaridade de cosseno. Essa medida tem como foco o ângulo existente entre os vetores de características que representam cada documento (Baeza-Yates e Ribeiro-Neto, 1999). Quando o resultado da similaridade do cosseno é 0 significa que não existem termos comuns entre eles (os vetores são ortogonais), e quando o resultado é 1, que os vetores são idênticos. A similaridade do cosseno entre os vetores v_1 e v_2 é calculada por meio da Equação 4.10, que representa o produto interno entre os elementos do vetor.

$$\text{Cosseno}_{\vec{v}_1, \vec{v}_2} = \frac{\sum_{i=1}^n w_{qi} w_{di}}{\sqrt{\sum_{i=1}^n w_{qi}^2} * \sqrt{\sum_{i=1}^n w_{di}^2}} \quad (4.10)$$

Esses valores de similaridade de cosseno são calculados para todos os vetores de características correspondentes aos documentos que representam cada conceito da ontologia. A partir desses valores, elabora-se uma matriz de similaridade entre os conceitos da ontologia, a qual é utilizada no refinamento dos graus de relacionamento entre os mesmos.

- Refinamento dos Graus de Relacionamento entre Conceitos: Após a obtenção dos graus de relacionamento existentes entre os diferentes conceitos da ontologia, o próximo passo é o seu refinamento. Essa atividade consiste na depuração de alguns relacionamentos de similaridade considerados irrelevantes de acordo com o valor atribuído ao seu grau de relacionamento. Essa depuração é feita considerando-se um valor limite, a fim de decidir quais relacionamentos possuem valores aceitáveis para o domínio escolhido. O valor limite é uma constante estabelecida de acordo com a estrutura do documento. Como se trata de um documento do tipo atributo-valor, a parte que corresponde à representação dos atributos no texto é similar em todo o conjunto, incrementando o valor de similaridade. Esse incremento deve ser controlado por meio da depuração de dados de acordo com o valor limite definido.

- Modelagem dos Conceitos e dos Relacionamentos: A última atividade a ser desenvolvida é a representação dos conceitos e relacionamentos existentes na ontologia. Considerando que os valores obtidos para os relacionamentos não pertencem ao conjunto booleano de dados, quer dizer que são valores vagos ou imprecisos, utilizou-se uma ontologia de alto nível (Yaguinuma et al., 2010), (Cerri, 2010). Tal ontologia modela classes e relacionamentos difusos para serem herdados ou instanciados pelas ontologias específicas de domínio, de modo que essas sejam capazes de representar informações imprecisas em um contexto específico. Na Figura 4.7 é ilustrado como a meta-ontologia permite representar os relacionamentos difusos de forma abstrata.

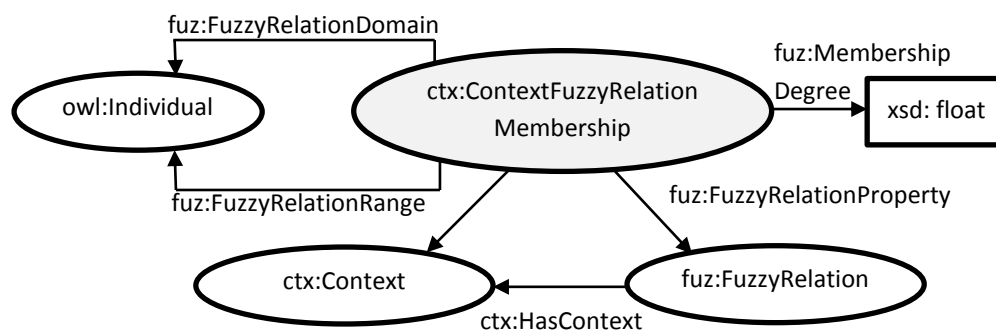


Figura 4.7: Representação Abstrata de um Relacionamento Binário Difuso

Além disso, na Figura 4.7 é ilustrado como a meta-ontologia representa relacionamentos difusos binários por meio da classe *ContextFuzzyRelationMembership*, responsável por atribuir um grau de pertinência (*MembershipDegree*) ao relacionamento difuso (*FuzzyRelationProperty*) entre duas instâncias (*FuzzyRelationDomain* e *FuzzyRelationRange*) para um determinado contexto (*Context*). O relacionamento *FuzzyRelationProperty* é uma propriedade de anotação, pois associa instâncias de *ContextFuzzyRelationMembership* a relacionamentos de OWL, além de associá-los a um determinado contexto (*HasContext*).

4.1.2 Engenharia do Sistema de Recomendação

A engenharia do Sistema de Recomendação pode ser considerada uma instância da Engenharia de Aplicação comumente utilizada de forma generalizada quando se trata de construir aplicações com foco na reutilização de

software. Assim, pode-se deduzir que a engenharia do Sistema de Recomendação dedica-se ao estudo das melhores técnicas, processos e métodos para a construção de aplicações de recomendação, tendo como base o reuso de artefatos. Aqui, os artefatos, ontologias, produzidos na engenharia de ontologia são reutilizados para o desenvolvimento das aplicações do domínio do problema considerado (Griss et al., 1998). No diagrama SADT da Figura 4.8 são ilustradas as etapas da engenharia do Sistema de Recomendação: Analisar, Projetar, Implementar e Testar. Essas atividades correspondem às disciplinas já conhecidas nos processos de desenvolvimento de software, porém adaptadas para atender ao reuso das ontologias produzidas na engenharia de ontologia.

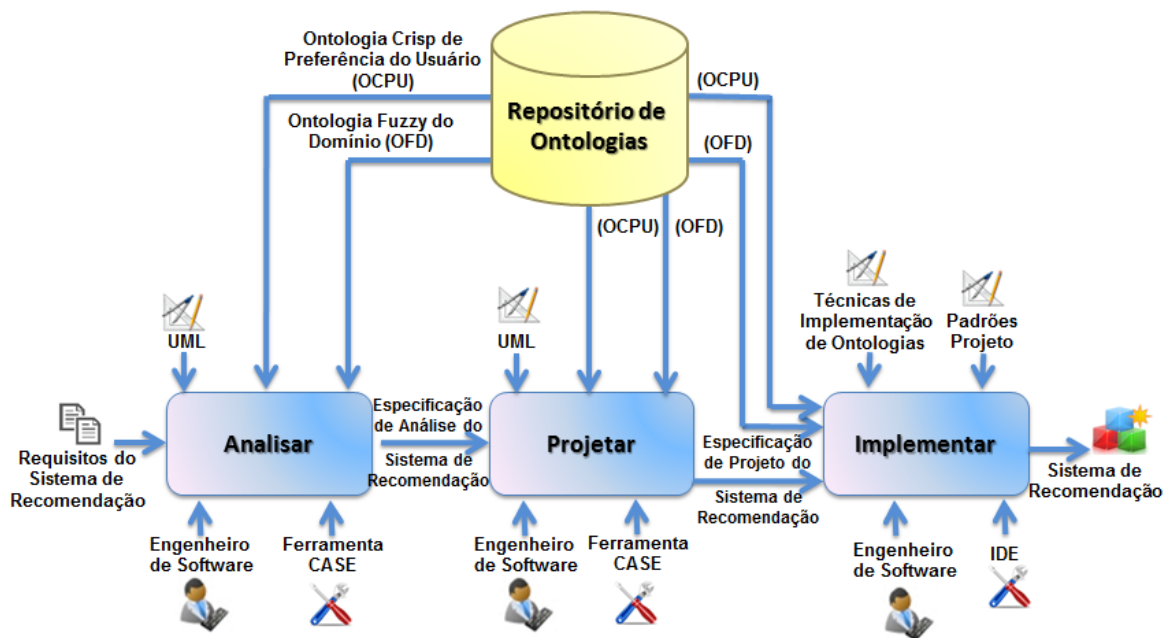


Figura 4.8: Visão Geral da Engenharia do Sistema de Recomendação

Durante a fase de análise, os requisitos da aplicação são analisados e especificados por meio da utilização de técnicas UML. Já durante a fase de projeto as especificações da aplicação são refinadas, considerando as tecnologias, plataformas de hardware e de software e outras decisões do projeto. Finalmente, durante a última fase a aplicação é implementada e testada. Uma ilustração mais detalhada do desenvolvimento de cada uma das fases da engenharia do sistema de recomendação será apresentada a seguir. Um estudo de caso foi conduzido a fim de desenvolver uma aplicação de recomendação no domínio dos dispositivos móveis.

4.2 Estudo de Caso: Um Sistema de Recomendação Conforme a Abordagem ObRA

Com o crescente volume de informações, os Sistemas de Recomendação têm por objetivo reduzir a sobrecarga de informação por meio da seleção de conteúdo baseado em preferências de usuário. Esses sistemas são classificados em três categorias, dependendo da abordagem utilizada para a recomendação: a) abordagem baseada em conteúdo, na qual são recomendados itens similares àqueles que o usuário demonstrou preferência no passado; b) abordagem colaborativa, na qual são recomendados itens escolhidos por pessoas com preferências similares às do usuário; e c) abordagem híbrida, a qual procura combinar as duas anteriores ([Adomavicius e Tuzhilin, 2005b](#)).

A fim de avaliar a aplicabilidade da abordagem proposta, apresenta-se um estudo de caso conduzido em um ambiente de comércio de dispositivos móveis. O estudo de caso foi executado segundo as etapas da metodologia apresentada neste capítulo: (1) Engenharia de Ontologia, que detalha o processo de fuzzificação de uma ontologia crisp, assim como a elaboração de uma ontologia crisp de preferências de usuário, e (2) Engenharia da Aplicação, que detalha e avalia os processos de propagação do interesse do usuário.

4.2.1 Engenharia de Ontologia: Ontologia *Fuzzy* do Domínio de Dispositivos Móveis e Ontologia Crisp de Preferências de Usuário

A adaptação UPFON foi elaborada com o intuito de permitir que o engenheiro de ontologia gere uma ontologia *fuzzy*, assim como permitir fuzzificar uma ontologia crisp já existente. Para verificar a viabilidade da atividade de fuzzificação de uma ontologia existente, utilizou-se uma adaptação da ontologia proposta em [Zhu et al. \(2010\)](#), com foco no domínio de comércio de dispositivos móveis. Na Figura 4.9 é ilustrado um trecho da ontologia utilizada neste estudo.

O processo de fuzzificação de uma ontologia crisp é definido pelo fluxo de projeto presente na adaptação UPFON, embora sejam necessárias a realização das atividades de definição dos contextos e a aquisição de recursos específicos do domínio, pertencentes, respectivamente, aos fluxos de requisitos e aos de análise.

A determinação do contexto é realizada por meio da análise do domínio, já

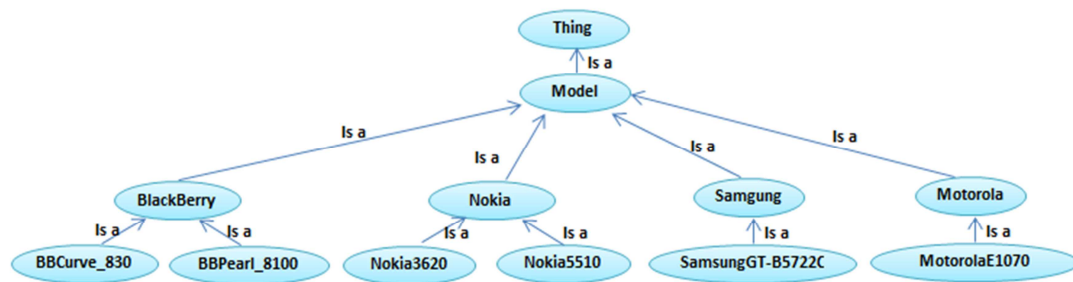


Figura 4.9: Trecho da ontologia no domínio de dispositivos móveis, utilizada para o estudo

que são os seus fatores que colaboram com a definição dos contextos sobre os quais se aplica o relacionamento de similaridade utilizado na abordagem proposta. No estudo de caso, definiu-se como contexto, o conjunto de características pertencentes a cada dispositivo. Assim, a aquisição de recursos específicos foi feita considerando-se o contexto definido e por meio da coleta de documentos com especificações técnicas das características dos dispositivos móveis. Esses documentos foram obtidos a partir do repositório público de dados XML, chamado *Wireless Universal Resource Files (WURFL²)*. Na Figura 4.10, encontra-se um trecho do repositório WURFL, no qual as características associadas aos dispositivos podem ser identificadas.

```

<device id="samsung_x830_ver1" user_agent="SEC-SGHX830/1.0 NetFront/3.2 Profile/MIDP-2.0
  <group id="product_info">
    <capability name="mobile_browser" value="Access Netfront"/>
    <capability name="mobile_browser_version" value="3.2"/>
    <capability name="uaprof" value="http://wap.samsungmobile.com/uaprof/x830_10.xml"/>
    <capability name="model_name" value="SGH-X830"/>
    <capability name="brand_name" value="Samsung"/>
  </group>
  <group id="display">
    <capability name="columns" value="6"/>
    <capability name="physical_screen_height" value="32"/>
    <capability name="physical_screen_width" value="19"/>
    <capability name="max_image_width" value="120"/>
    <capability name="rows" value="17"/>
    <capability name="resolution_width" value="128"/>
    <capability name="resolution_height" value="220"/>
    <capability name="max_image_height" value="210"/>
  </group>

```

Figura 4.10: Trecho do repositório WURFL, indicando as características correspondentes ao dispositivo Samsung_X830

Uma vez determinado o contexto no qual se realizará o relacionamento e selecionado os recursos específicos do domínio, se segue com a realização das

²<http://wurfl.sourceforge.net/>

atividades do Fluxo de Fuzzificação. Na primeira atividade do fluxo, os documentos obtidos são pré-processados com o intuito de obter uma representação vetorial de cada documento, permitindo, assim, elaborar uma matriz do tipo atributo-valor com os dispositivos e as suas características. Essa representação pode ser obtida por meio da utilização de ferramentas de Mineração de Texto, como a ferramenta PreText (Soares et al., 2008), a qual, mediante o cálculo da frequência de cada um dos termos presentes nos documentos, permite a elaboração de uma matriz atributo-valor para esse conjunto de documentos.

Uma vez obtida essa matriz, segue-se com a determinação dos graus de relacionamento dos conceitos. Vale ressaltar que, para isso, é necessário realizar o cálculo de similaridade entre os documentos. Esse cálculo servirá como indicador para determinar a similaridade entre os conceitos da ontologia. Isso acontece devido ao fato de que cada conceito (dispositivo) é representado por um documento. O cálculo de similaridade é realizado por meio da equação de similaridade do cosseno, a qual calcula a diferença existente entre dois vetores, com base no cosseno do ângulo entre ambos os vetores. Esse tipo de cálculo é bastante utilizado em áreas de Mineração de Texto e Estatística, as quais possuem diversas ferramentas de apoio para esse cálculo. Para este estudo de caso, o cálculo da similaridade foi executado com a ajuda da ferramenta RGui (Faria, 2011) utilizando o pacote TM, o qual é especializado em técnicas de Mineração de Textos. O resultado desse cálculo é uma matriz de similaridade onde todos os conceitos identificados na matriz atributo-valor possuem um grau de similaridade. Na Tabela 4.1 é mostrado um trecho da matriz obtida no processamento.

Tabela 4.1: Trecho da Matriz de Similaridade obtida na atividade de determinação do grau de relacionamento dos conceitos

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
(1) blackberry9520	0,000	0,072	0,050	0,060	0,124	0,072	0,078	0,062
(2) blackberry0857	0,072	0,000	0,060	0,080	0,138	0,100	0,070	0,066
(3) blackberry8220	0,050	0,060	0,000	0,026	0,106	0,064	0,066	0,060
(4) blackberry8120	0,060	0,080	0,026	0,000	0,094	0,052	0,072	0,058
(5) blackberry8110	0,124	0,138	0,106	0,094	0,000	0,120	0,114	0,104
(6) blackberry8900	0,072	0,100	0,064	0,052	0,120	0,000	0,074	0,074
(7) blackberry9100	0,078	0,070	0,066	0,072	0,114	0,074	0,000	0,052
(8) blackberry9700	0,062	0,066	0,060	0,058	0,104	0,074	0,052	0,000

Após os valores de similaridade serem obtidos, procede-se com o refinamento dos graus de relacionamento entre conceitos. Nessa atividade, é estabelecido um limite de similaridade, o qual permite que os relacionamentos com valores pouco significativos sejam desconsiderados. Para o estudo de caso, considerou-se um limite de 80%, tendo em vista a similitude existente na estrutura dos documentos. Uma vez concluída a etapa de refinamento dos graus de similaridade dos conceitos, a última atividade a se desenvolver é a de modelagem hierárquica dos relacionamentos de similaridade. Como apresentado no capítulo 4, a modelagem hierárquica é realizada utilizando-se a abordagem proposta por Yaguinuma et al. (2010); Cerri (2010), uma vez que essa permite representar de forma abstrata o conceito de relacionamento *fuzzy* presente na ontologia. Na Figura 4.11, ilustra-se a representação abstrata do relacionamento de similaridade entre dois conceitos da ontologia.

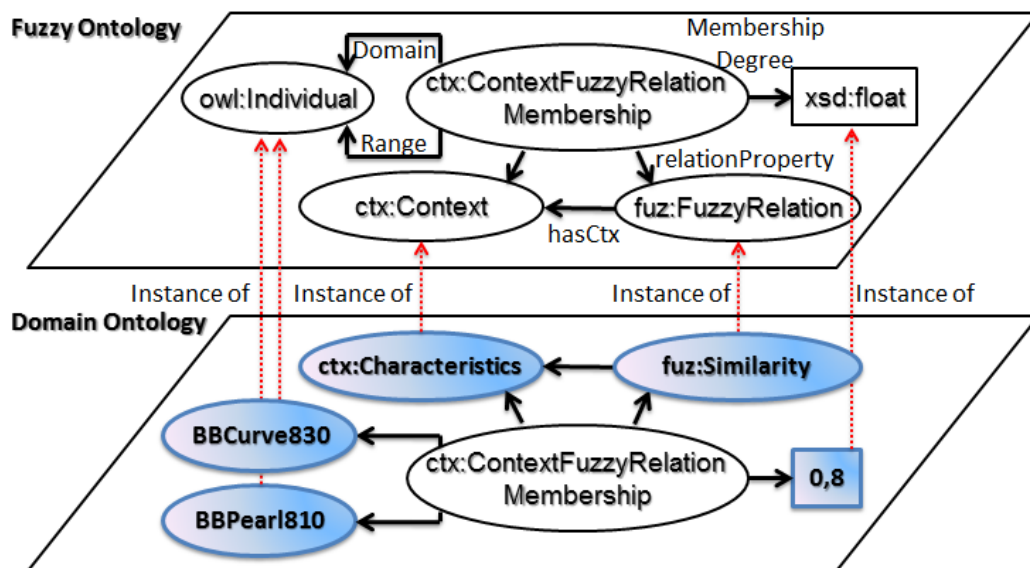


Figura 4.11: Representação abstrata do relacionamento de similaridade utilizando a meta-ontologia *fuzzy* proposta por Yaguinuma et al. (2010); Cerri (2010)

Considerando a ontologia do domínio e os relacionamentos *fuzzy* `SimilarTo`, as sub-propriedade de `FuzzyRelation`, e as instâncias de `ContextFuzzyRelationMembership` que relacionam as instâncias da ontologia ao valor do grau do relacionamento *fuzzy* são modeladas. Na Figura 4.11, apresenta-se o relacionamento *fuzzy* `similarTo` entre os dispositivos “BlackBerry_Pearl 8100” e “BlackBerry_Curve 830” com grau de similaridade = 0.8. Uma vez calculados os valores de similaridade e modelados os relacionamentos difusos

faz-se a codificação da ontologia. A linguagem de representação de ontologias OWL é utilizada para esse fim. Na Figura 4.12, apresenta-se um trecho do código da ontologia *fuzzy* de domínio.

```
<!-- http://www.owl-ontologies.com/ontFuzzy.owl#BlackBerry_8110 -->

<owl:NamedIndividual rdf:about='&ontFuzzy;BlackBerry_8110'>
  <rdf:type rdf:resource='&ontFuzzy;RIM' />
  <fuz:similarTo_directly rdf:resource='&ontFuzzy;BlackBerry_857' />
  <owl:Axiom>
    <rdfs:comment>BlackBerry_8110-BlackBerry_857</rdfs:comment>
    <fuz:fuzzySimilarityDegree>0.861723447</fuz:fuzzySimilarityDegree>
    <owl:annotatedProperty rdf:resource='&fuz:similarTo_directly' />
    <owl:annotatedSource rdf:resource='&ontFuzzy;BlackBerry_8110' />
    <owl:annotatedTarget rdf:resource='&ontFuzzy;BlackBerry_857' />
  </owl:Axiom>
```

Figura 4.12: Representação formal da ontologia *fuzzy* de domínio

Como se pode observar na Figura 4.12, a representação formal da ontologia baseia-se nas abstrações propostas na meta-ontologia *fuzzy* (Yaguinuma et al., 2010).

As atividades de teste e de verificação de um processo de fuzzificação tem como foco a validação dos valores obtidos nos relacionamentos de similaridade. Essa validação é realizada por um especialista do domínio específico, de forma a determinar se os valores são condizentes com a realidade. Outra forma de validar os resultados da fuzzificação da ontologia é por meio da implementação de uma aplicação que utilize a ontologia *fuzzy* com o objetivo de recomendar informações relevantes ao usuário. A análise dos resultados obtidos com a aplicação permite determinar se esses são condizentes com os que o usuário procura. A fim de implementar essa aplicação, é necessário manter as preferências de usuário armazenadas e atualizadas com o auxílio de uma ontologia crisp de preferência de usuário.

Essa ontologia é modelada como uma instância da ontologia de domínio, com a diferença de que cada conceito tem um valor de interesse de usuário associado. Nas Figuras 4.13 e 4.14 são ilustradas as representações gráfica e formal de uma parte da ontologia crisp de preferência de usuário. Nessas representações, pode-se destacar o atributo “interesse de usuário” (*userInterest*), que inicialmente se encontra vazio para todos os usuários.

Uma vez elaborada a ontologia crisp de preferência de usuário, a próxima etapa é a validação da utilização de ontologias *fuzzy* de domínio juntamente com a propagação de interesses de usuário. Essa validação visa determinar

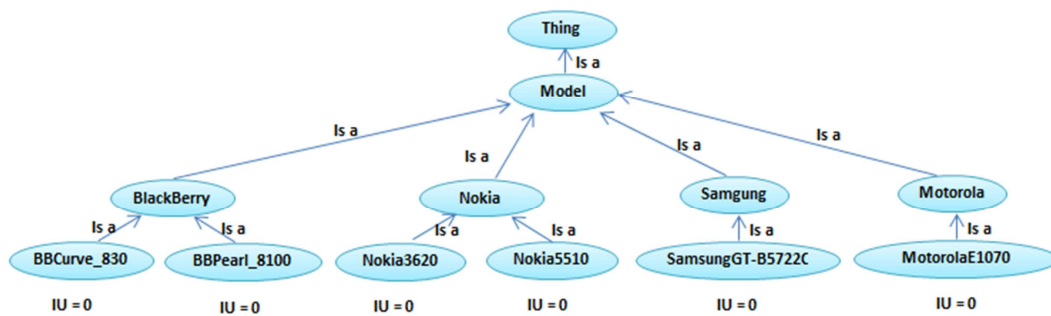


Figura 4.13: Representação gráfica de uma parte da ontologia crisp de preferência de usuário

```

<owl:NamedIndividual rdf:about='&ontFuzzy;1110'>
  <rdf:type rdf:resource='&ontFuzzy;Nokia' />
  <rdfs:label>1110</rdfs:label>
  <userInterest>1.7463</userInterest>
  <conceptWeight>2</conceptWeight>
  <conceptData>21/04/2012</conceptData>
  <fuz:similarTo_directly rdf:resource='&ontFuzzy;CDM-180' />
  <fuz:similarTo_directly rdf:resource='&ontFuzzy;BlackBerry_9800' />
  <fuz:similarTo_directly rdf:resource='&ontFuzzy;T65' />
  <fuz:similarTo_directly rdf:resource='&ontFuzzy;i919' />
  <fuz:similarTo_directly rdf:resource='&ontFuzzy;i921' />

```

Figura 4.14: Representação formal de uma parte da ontologia crisp de preferência de usuário

se o uso das ontologias facilita o processo de filtragem de informação. Para isso desenvolveu-se uma aplicação de recomendação, cujo detalhamento é apresentado na etapa de engenharia de aplicação, a seguir.

4.2.2 Engenharia do Sistema de Recomendação

A etapa de Engenharia de Aplicação iniciou-se com a atividade de análise. Nessa atividade, o Engenheiro de Software tendo por referência os requisitos do Sistema de Recomendação, especifica-os em UML utilizando uma ferramenta CASE. As ontologias, Fuzzy e Crisp, representadas em OWL, orientam a elaboração dessas especificações. Na Figura 4.15 é ilustrada uma parte do modelo de classes do SR. As classes Usuario, PreferênciaUsuario, Recomendação e Dispositivo foram especificadas a partir da ontologias crisp de preferência do usuário e tem como objetivo acessar as ontologias para obter os dados que permitam calcular os valores de interesse real e propagado para cada dispositivo para um determinado usuário. A classe DispositivoSimilar

está diretamente relacionada com a ontologia fuzzy de domínio e tem como objetivo acessar aos relacionamentos de similaridade existentes dentro da ontologia.

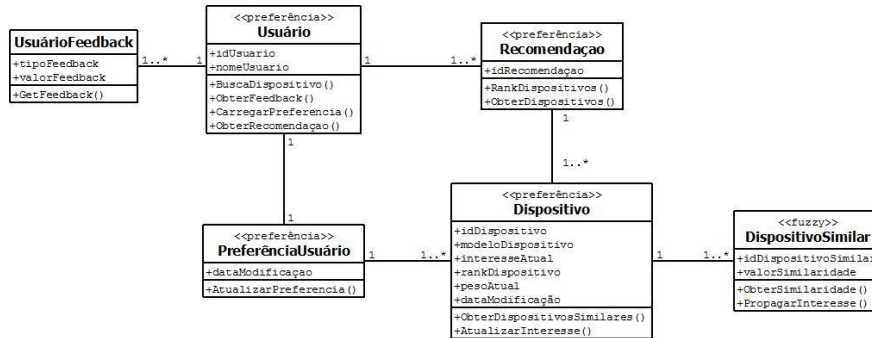


Figura 4.15: Parte do Modelo de Classes como especificação de Análise

Após a análise, realiza-se a atividade de projeto. Nessa atividade projetar as especificações da análise são refinadas considerando as decisões de projeto, como a utilização do padrão MVC resultando na divisão de responsabilidade entre classes. Por exemplo, no modelo da Figura 4.16 observa-se como a classe `IPanelRecomendação`, pertencente a camada de vista, permite visualizar a recomendação fornecida pela classe de modelo `Recomendação` por meio da classe de controle `RecomenderAction` que é a encarregada de disparar a execução das funções que acessam as ontologias. Outra decisão de projeto a considerar é a definição das plataformas de implementação, que para este projeto foram a plataforma *DotNet* e o *framework DotNetRDF*.

Nessa atividade é definida também a modelagem da arquitetura da aplicação, como ilustrado na Figura 4.17. Pode-se observar na figura que a comunicação entre a aplicação e o Servidor Web é realizada via HTTP e utilizando um Serviço Web baseado em REST, que permite uma comunicação mais rápida. Os dados da ontologia *fuzzy* de domínio, assim como das ontologias de usuário encontram-se armazenados em um Servidor de Ontologias e são acessados diretamente pelo servidor Web. Os resultados fornecidos pelo Servidor de Ontologias são processados no Servidor Web e retornados à aplicação para serem formatados e mostrados ao usuário.

Uma vez definida a arquitetura e modelados os conceitos provenientes da atividade de análise, a próxima atividade a ser realizada é a implementação da aplicação. Nessa atividade, reutilizam-se as ontologias geradas na engenharia de ontologia para implementar as especificações de projeto. O reuso

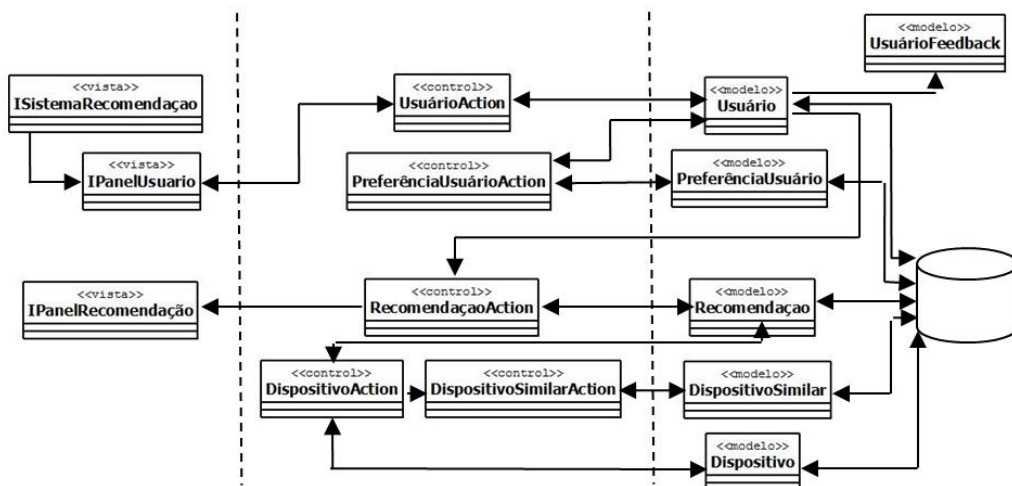


Figura 4.16: Representação Modelo Vista Controlador como parte da especificação de Projeto

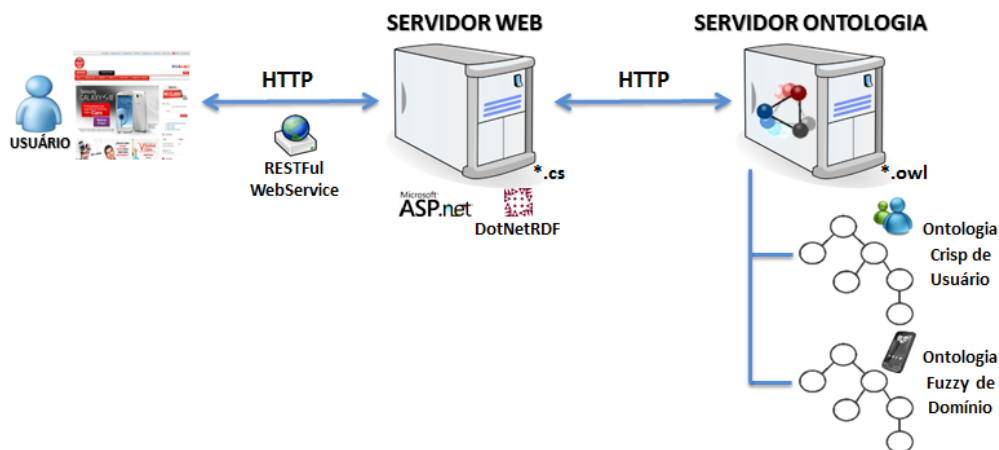


Figura 4.17: Visão geral da arquitetura da aplicação

ocorre conforme ilustrado na Figura 4.18. A ontologia crisp de preferência de usuário é instanciada utilizando um objeto da classe Graph e associando-o ao arquivo .owl da ontologia. Para acessar aos dados da ontologia, é utilizado um objeto da classe SparqlResultSet. Esse objeto permite realizar uma consulta à ontologia. Essa consulta é feita por meio do sparql que utiliza o reasoner da ontologia para inferir dados e relacionamentos explícitos presentes na ontologia.

Nessa atividade também é realizada a propagação do interesse do usuário baseado na similaridade existente entre os conceitos. Uma vez identificados

```

//Load Ontology
Graph grOntology = new Graph();
FileLoader.Load(grOntology, "C:\\Users\\Arturo\\Documents\\Visual Studio 2010\\OntoUsuario.owl");

Object objResult = tsStore.ExecuteQuery(strQuery);
if (objResult is SparqlResultSet)
{
    SparqlResultSet rset = (SparqlResultSet)objResult; //Convert to SparqlResultSet
    foreach (SparqlResult srResult in rset) //Read TripleStore Values
    {
        String strNewString;
        strNewString = ProcessString(srResult.ToString());
        strNewString = SplitString(strNewString, "Target");
        strList =strNewString;
    }
}

```

Figura 4.18: Trecho do código para acesso aos dados da ontologia

os conceitos de interesse do usuário, o valor calculado tem que ser propagado para os conceitos diretamente relacionados a ele. Esse valor de interesse propagado é obtido baseado nas fórmulas adaptadas do trabalho de (Cena et al., 2011), para considerar o relacionamento de similaridade $d(N,M)$ como fator de propagação, como parte da equação 1 mostrada a seguir:

$$I_p(Nokia5510) = \frac{e^{(-k * d(MotorolaE1070, Nokia5510))}}{1 + \log(1 + n(Nokia5510))} * I_s(MotorolaE1070) = 2.0642 \quad (4.11)$$

A Figura 4.19 mostra o trecho de código que representa o calculo do interesse propagado sobre o conceito M. Para isso, na equação 4.11 é considerado um coeficiente k que permita definir qual o impacto da propagação sobre esse nó. O valor $d(N,M)$ representa o grau de similaridade existente entre os conceitos N e M, e é obtido a partir da ontologia fuzzy do domínio. Os dados da ontologia fuzzy do domínio estão representadas numa estrutura structConceptInstance com a finalidade de ter acesso a cada um dos atributos existentes na ontologia. A função $n(M)$ representa o numero de ações realizadas sobre o conceito M, e o valor $I_s(N)$ representa o interesse atual sobre o conceito N.

Durante a implementação, é necessário considerar que cada interação do usuário com a aplicação significa a modificação dos valores associados aos interesses do usuário. Essa modificação está diretamente relacionada ao tipo de ação realizada pelo usuário (*feedback*), sendo que cada tipo possui um valor de influência associado preestabelecido. No estudo de caso proposto, foram considerados dois tipos de feedback:

- Acessar a página relacionada ao dispositivo, com valor 5;

```

strSimilarConcept = structConceptsInstance.strTarget.ToString();
strSimilarityValue = structConceptsInstance.strSimilarity.ToString();

dblConceptWeight = Convert.ToDouble(objQuery.GetConceptWeight(strSimilarConcept));

dblValorExponential = dblAuxFormula * (1-(Convert.ToDouble(strSimilarValue))) * (-1);
dblValorExponential = System.Math.Pow(dblAuxFormulaNeperiano, dblValorExponential);
dblValorLogarithm = 1 + System.Math.Log(1 + dblConceptWeight);

// IPr_actual(n) = (e ^ -k*simConcepts/(1 + Log(1+ weightConcepts)))*(userInterest)
dblActualPropagatedUserInterestValue = (dblValorExponential / dblValorLogarithm) * dblUserInterest;

return dblActualPropagatedUserInterestValue;

```

Figura 4.19: Trecho do código para a propagação do interesse

- Adicionar o dispositivo como favorito, com valor 7.

A fim de ilustrar o processo de atualização de preferências realizado, supõe-se que o usuário acesse a página associada ao conceito “MotorolaE1070”, com valor do *feedback* é 2. A primeira ação a ser executada é a obtenção dos conceitos similares baseados na ontologia *fuzzy* de domínio. Os conceitos similares relevantes obtidos são “Nokia 5510” e “SamsungGT-B5722C”, com um valor de similaridade de 0.9 e 0.8, respectivamente. A próxima etapa consiste em calcular o valor do interesse do usuário sobre o conceito (nó) inicial $I(MotorolaE1070)$ e, para isso, é necessário obter o interesse atual do nó $I_s(MotorolaE1070)$, conforme cálculo apresentado na equação 4.12:

$$I_s(MotorolaE1070) = \frac{l(MotorolaE1070) + 1}{max + 1} * f(MotorolaE1070)^b = 2.2361 \quad (4.12)$$

Como o valor anterior (inicial) do interesse do usuário sobre o nó inicial é 0, o interesse final para esse nó se equipara ao interesse atual do nó $I_s(MotorolaE1070)$. Uma vez obtido o interesse final, esse precisa ser propagado para os demais nós que representem conceitos similares. O valor de interesse do nó similar $I(Nokia5510)$ é obtido por meio do cálculo da propagação do interesse $I_s(Nokia5510)$ sobre esse nó, acrescido do valor anterior correspondente sobre esse mesmo nó. O interesse propagado do nó $I_p(Nokia5510)$ calculado de acordo com a equação 4.13 é 2.0642.

$$I_p(Nokia5510) = \frac{e^{(-k * d(MotorolaE1070, Nokia5510))}}{1 + \log(1 + n(Nokia5510))} * I_s(MotorolaE1070) = 2.0642 \quad (4.13)$$

Considerando que o valor anterior de interesse do nó é 0, o valor atual de

interesse é igual ao valor de interesse propagado do nó $I_p(Nokia5510)$. Seguindo o mesmo cálculo, o valor obtido para o nó propagado $I_p(SamsungGT - B5722C)$ é 1.9054. Uma vez que os valores de interesse de usuário foram calculados, deve-se considerar que os valores do número de ações sobre o nó $I(MotorolaE1070)$ foram incrementados.

As ações seguintes realizadas pelos usuários utilizam os valores do exemplo anterior como valores anteriores de interesse de usuário. Por exemplo, se o usuário acessa a página associada ao conceito $Nokia5510$, tal conceito já possui um valor de interesse anterior $I_o(Nokia5510)$ de 2.0642. O cálculo para obter o seu novo valor de interesse é apresentado na equação 4.14.

$$I(Nokia5510) = I_o(Nokia5510) + I_s(Nokia5510) = 4.3003 \quad (4.14)$$

Propagando-se o valor de interesse calculado para $I(Nokia5510)$ sobre os nós com os quais existe uma relação de similaridade, obtém-se um valor propagado para $I(Nokia3620)$ de 3.9697 e para $I(MotorolaE1070)$ de 5.2873. Os valores finais dos interesses de usuário, assim como o número de ações realizadas por ele são mostrados na Tabela F.1.

Tabela 4.2: Resultados dos interesses do usuário

Conceitos		Início		Consulta 1		Consulta 2	
		Pref.	n(M)	Pref.	n(M)	Pref.	n(M)
1	MotorolaE1070	0	0	2.2361	1	5.2873	1
2	Nokia5510	0	0	2.0642	0	4.3003	1
3	SamsungGTB5722C	0	0	1.9054	0	1.9054	0
4	Nokia3620	0	0	0	0	3.9697	0

Nessa Tabela é possível observar que o valor do interesse sobre um conceito é incrementado de acordo com o número de vezes que o conceito é utilizado na propagação. Esses interesses finais são utilizados para estabelecer o ranqueamento das preferências do usuário que permitiram oferecer uma recomendação final.

4.3 Avaliação da Estratégia de Propagação de preferência do usuário

Nesta seção são descreve-se uma série de testes realizados a fim de avaliar a influência da propagação dos interesses dos usuários nos processos de recomendação. Com o objetivo de calcular a influência da propagação nos resultados, cinco medidas de avaliação foram consideradas: (1) valor do erro absoluto (MAE), (2) precisão, (3) cobertura, (4) *F-score* e (5) correlação de *Spearman*. Para isso, é necessário realizar a coleta dos interesses dos usuários sobre um conjunto determinado de dispositivos. Esses dados são comparados aos dados obtidos pelo Sistema de Recomendação, permitindo, assim, realizar o cálculo das medidas de avaliação estabelecidas. Na Tabela 4.3 é mostrada a comparação entre os valores fornecidos pelo usuário 01 em comparação aos valores obtidos por meio do Sistema de Recomendação.

Tabela 4.3: Dispositivos relevantes ranqueados de acordo com as preferências do usuário e a abordagem de recomendação sem considerar a propagação do interesse

	Marcação Usuário	Marcação SR sem propagação
Posição 1	Dispositivo 8	Dispositivo 8
Posição 2	Dispositivo 10	Dispositivo 2
Posição 3	Dispositivo 2	Dispositivo 10
Posição 4	Dispositivo 5	Dispositivo 9
Posição 5	Dispositivo 7	Dispositivo 3

Nessa Tabela pode-se observar que os valores recomendados pelo sistema não são exatos em relação aos fornecidos pelo usuário. Na Tabela 4.4 são mostrados os valores calculados para cada uma das medidas de avaliação, considerando os dados mostrados na Tabela 4.3.

Com a finalidade de avaliar a influência das mudanças nas preferências dos usuários no Sistema de Recomendação, foram realizadas duas alterações, mediante interações entre o usuário e o sistema, as quais permitiram modificar as preferências do usuário. Essas modificações resultaram em recomendações mais condizentes com as preferências do usuário, como apresentado na Tabela 4.5.

Na Tabela 4.6 são mostrados os valores calculados para cada uma das medidas de avaliação, considerando os dados mostrados na Tabela 4.5.

Tabela 4.4: Cálculo das medidas das métricas de precisão da abordagem de recomendação sem propagação para o conjunto de Dispositivos1

	MAE	Precisão	Cobertura	F-Score	Spearman
Usuário 1	0.8	0.71	1.00	0.83	0.75
Usuário 2	2.2	0.43	0.60	0.50	0.07
Usuário 3	2.2	0.43	0.60	0.50	0.14
Usuário 4	0.4	0.57	0.80	0.67	0.39
Usuário 5	1.6	0.57	0.80	0.67	-0.07
Usuário 6	2.2	0.57	0.80	0.67	-0.14
Usuário 7	1.8	0.57	0.80	0.67	-0.04
Usuário 8	1.8	0.71	1.00	0.83	0.32
Usuário 9	1.8	0.57	0.80	0.67	-0.04
Usuário 10	1.0	0.43	0.60	0.50	0.32

Tabela 4.5: Dispositivos relevantes ranqueados de acordo com as preferências do usuário e a abordagem de recomendação considerando a propagação do interesse

	Marcação Usuário	Marcação SR com propagação
Posição 1	Dispositivo 8	Dispositivo 2
Posição 2	Dispositivo 10	Dispositivo 10
Posição 3	Dispositivo 2	Dispositivo 3
Posição 4	Dispositivo 5	Dispositivo 8
Posição 5	Dispositivo 7	Dispositivo 7

Comparando-se os resultados das medições utilizadas para a avaliação da precisão dos resultados, pode-se concluir que a propagação dos interesses de usuário por meio de interações entre o usuário e o sistema contribuiu de forma positiva no processo de recomendação. Essa contribuição envolve a redução do valor de erro absoluto, o incremento no F-score e na correlação existente entre o ranqueamento de ambas as recomendações.

4.4 Considerações Finais

Este capítulo apresentou a abordagem para apoio à construção de um sistema de recomendação baseado em ontologias, desenvolvida neste trabalho.

Tabela 4.6: Cálculo das medidas das métricas de precisão da abordagem de recomendação com propagação para o conjunto de Dispositivos1

	MAE	Precisão	Cobertura	F-Score	Spearman
Usuário 1	1.2	0.71	1.00	0.83	-0.21
Usuário 2	1.4	0.43	0.60	0.50	-0.29
Usuário 3	1.8	0.43	0.60	0.50	-0.32
Usuário 4	1.8	0.57	0.80	0.67	-0.25
Usuário 5	2.0	0.57	0.80	0.67	0.68
Usuário 6	2.0	0.57	0.80	0.67	0.29
Usuário 7	1.6	0.57	0.80	0.67	-0.64
Usuário 8	1.4	0.57	0.80	0.67	0.14
Usuário 9	1.8	0.57	0.80	0.67	-0.54
Usuário 10	1.2	0.57	0.80	0.67	0.00

Essa abordagem baseia-se na utilização de uma ontologia *fuzzy* de domínio, que permite estabelecer relacionamentos de similaridade entre conceitos, e uma ontologia *crisp* de preferências de usuário, que permite determinar os valores possível de interesse por parte do usuário, baseado nesses relacionamentos.

A abordagem é dividida em duas etapas: (1) Engenharia de Ontologia, que engloba as atividades para a elaboração e a fuzzificação de uma ontologia de domínio mediante a utilização da adaptação UPFON, assim como a elaboração de ontologias de usuário que permitam armazenar as suas preferências; e (2) Engenharia de Sistema de Recomendação, na qual as ontologias produzidas na Engenharia de Ontologia são utilizadas como apoio ao desenvolvimento de uma aplicação de recomendação. Essa etapa engloba as atividades de Análise, Projeto, Implementação e Testes do ciclo de vida de desenvolvimento dos processos de engenharia de software convencionais.

O trabalho proposto pode ser comparado desde dois pontos de vista, o primeiro, no referente às técnicas existentes para a fuzzificação de ontologias e, o segundo, no referente à utilização de ontologias de preferência de usuário em sistemas de recomendação.

No referente à fuzzificação a diferença com os trabalhos correlatos se encontra na utilização de documentos de domínio como base para a determinação do grau de similaridade existente entre dois conceitos. Parry (2004) utilizou também um conjunto de documentos, mas o objetivo dele foi determinar

o grau de pertinência de um conceito com o documento para assim estabelecer uma hierarquia de termos. Já os outros trabalhos utilizam escores para avaliar cada conceito em relação ao conjunto de dados totais.

No referente à utilização de ontologias de preferências de usuário em sistemas de recomendação a diferença com os trabalhos correlatos se encontra na utilização de uma instancia da ontologia de domínio como ontologia de preferência de usuário e na propagação dessas preferências baseadas na similaridade existente entre conceitos. Alguns trabalhos, como por exemplo, o trabalho de [Sieg et al. \(2007\)](#) utiliza instancias da ontologia de domínio para representar a ontologia de preferencias de usuário, porem, não permitem realizar uma propagação desses interesses para outros conceitos relacionados. No trabalho de [Jiang e Tan \(2006\)](#) é realizada a propagação dos interesses, porém, essa propagação é feita utilizando a hierarquia e não os relacionamentos existentes entre os conceitos, por exemplo, o de similaridade abordado neste trabalho. Já outros trabalhos baseiam-se só na utilização do perfil do usuário como base para a ontologia de preferências de usuário, utilizando logs e históricos para a atualização do mesmo.

Além disso, foi apresentado um estudo de caso da aplicação da abordagem no domínio do comércio de dispositivos móveis. Nesse estudo de caso, realizou-se uma avaliação do comportamento dos resultados de um Sistema de Recomendação em relação ao incremento e propagação das preferências dos usuários.

A fim de testar se os resultados obtidos são consistentes com os interesses dos usuários realizou-se um experimento que permitiu medir de forma quantitativa, a correlação entre os resultados da aplicação de recomendação utilizando a abordagem baseada em ontologias *fuzzy* de domínio e em ontologias de preferências de usuário e uma aplicação de recomendação de itens utilizando uma abordagem baseada em agrupamento de dados. Esse processo de experimentação é apresentado no capítulo 5.

Avaliação Experimental da Abordagem ObRA

Neste capítulo são apresentados os estudos conduzidos com o objetivo de avaliar o impacto na precisão dos resultados obtidos com o uso de ontologias *fuzzy* e ontologias de usuário, como suporte à construção de Sistemas de Recomendação, em relação aos resultados obtidos com o uso de técnicas de agrupamento, como suporte à construção de Sistemas de Recomendação.

O capítulo está organizado da seguinte forma: a Seção 5.1 descreve o processo de desenvolvimento da experimentação. Na Seção 5.2.1, apresenta-se uma definição do experimento assim como os seus objetivos principais. Na Seção 5.2.2, apresenta-se o planejamento do experimento e as hipóteses utilizadas no processo. Em seguida, na Seção 5.2.3, são apresentadas e detalhadas as etapas da execução dos experimentos. Na Seção 5.2, são apresentadas as etapas do desenvolvimento do projeto. Finalmente, na Seção 5.3, apresenta-se a análise dos experimentos e a interpretação dos resultados.

5.1 Experimentação da Abordagem ObRA

O processo de experimentação da abordagem para a utilização de ontologias *fuzzy* e de ontologias de usuário como suporte para a construção de Sistemas de Recomendação seguiu todas as atividades experimentais definidas em Wohlin et al. (2000), Travassos et al. (2002) e Cirilo et al. (2011). O

experimento baseia-se em um estudo comparativo entre o uso da abordagem proposta, que utiliza ontologias *fuzzy* para a construção de Sistemas de Recomendação, e o uso de uma abordagem que utiliza agrupamento de dados com a mesma finalidade. O objetivo do experimento foi comparar os efeitos de ambas as abordagens em relação à precisão dos resultados obtidos. Nas seções seguintes são detalhadas as fases do desenvolvimento da experimentação.

5.2 Desenvolvimento da Experimentação

A fim de realizar o experimento, foram utilizadas duas aplicações: um conjunto de documentos e uma matriz de dados que representa os valores de similaridade entre os mesmos. Entre as aplicações utilizadas, tem-se uma aplicação de recomendação baseada em ontologias *fuzzy* de domínio e em ontologias de usuário, elaborada especialmente para a execução do experimento, e uma aplicação de agrupamento de documentos, que simula o comportamento de uma aplicação de recomendação baseada em agrupamento.

A tarefa dos participantes durante o experimento foi avaliar a similaridade dos dispositivos tendo como base um dispositivo de interesse. Para essa avaliação, foram utilizadas, como fator de similaridade, as características mais importantes de cada dispositivo, de acordo com as preferências de cada participante. Os dados coletados foram registrados em uma tabela de forma que os dispositivos mais similares ao dispositivo principal tivessem uma posição melhor no ranqueamento. Como o foco do estudo é a avaliação da precisão dos resultados recomendados pelas duas abordagens, os dados coletados da avaliação dos usuários foram utilizados no cálculo das métricas de medida de precisão utilizadas para avaliar os Sistemas de Recomendação. Essas métricas avaliam empiricamente o quão eficiente um Sistema de Recomendação é ao classificar itens para um determinado usuário, além de medir se o sistema pode ou não prever um valor exato de classificação para um item específico (Herlocker et al., 2000).

As fases experimentais do estudo são detalhadas nas subseções seguintes.

5.2.1 Definição do Experimento

Nessa fase, o experimento é definido em termos dos problemas e dos objetivos. Assim, a fase de definição estabelece a direção geral do experimento, o seu escopo, a base para a formulação das hipóteses e as direções iniciais para a avaliação da validade do estudo.

O objetivo desse experimento é **analisar** o uso de ontologias *fuzzy* para a representação de similaridade entre conceitos e o uso de ontologias de usuário para o armazenamento e a propagação de interesses como base dos Sistemas de Recomendação; **com o propósito** de avaliação; **com respeito** à eficiência em termos de medidas de precisão dos resultados obtidos dentro do conjunto de dados possíveis; **do ponto de vista** de especialistas do domínio; **no contexto** de usuários com interesse em dispositivos móveis.

O contexto de um experimento pode ser caracterizado em termos do número de participantes e dos objetos envolvidos no estudo experimental. Nesse experimento, considerando o envolvimento de vários participantes e o uso de duas abordagens para a recomendação de dados (abordagem baseada em ontologias e baseada em agrupamento), o contexto foi classificado como **estudo bloqueado de sujeito-objeto** (Wohlin et al., 2000). Isso significa que o estudo consistiu em diversos testes experimentais e, em cada um deles, coletou-se uma série de dados, a partir dos quais as medidas de precisão das duas abordagens foram avaliadas.

5.2.2 Planejamento do Experimento

Nessa fase, a fundamentação do experimento é estabelecida, a qual envolve as seguintes etapas:

1. **Seleção do Contexto:** O estudo foi realizado em um ambiente controlado, ou seja, com usuários de dispositivos móveis com conhecimento específico em computação.
2. **Formulação das Hipóteses:** Foram elaboradas três hipóteses para o experimento levando em conta o efeito da utilização da abordagem de recomendação no resultado. Para a formulação dessas hipóteses, foram consideradas as seguintes métricas:
 - *Métricas de Precisão da Predição:* determinam até que ponto um Sistema de Recomendação pode prever as avaliações dos usuários. Essa métrica é quantificada por meio da medição do erro absoluto (MAE).
 - *Métricas de Precisão da Classificação:* determinam o quão corretamente um Sistema de Recomendação é capaz de classificar os itens como interessantes ou não interessantes, de acordo com a classificação do usuário. Essa métrica é quantificada por meio da medição da precisão e da cobertura.

- *Métricas de Precisão do Ranqueamento*: determinam a capacidade de um Sistema de Recomendação para recomendar uma lista ordenada de resultados de acordo com a importância dos itens para o usuário. Essa métrica é quantificada por meio da medição do coeficiente de correlação de *Spearman*.

A hipótese nula e suas alternativas correspondentes são:

Hipótese Nula (H0W) : Em geral, não há diferença na precisão (P) dos resultados entre a abordagem de recomendação baseada em ontologias *fuzzy* de domínio e a baseada em agrupamento de dados.

$$H0W = P_{OntologiaFuzzy} = P_{Agrupamento} \Rightarrow PP_{OntologiaFuzzy} = PP_{Agrupamento}$$

$$\text{e } PC_{OntologiaFuzzy} = PC_{Agrupamento} \text{ e } PR_{OntologiaFuzzy} = PR_{Agrupamento}$$

Hipótese Alternativa (H1W) : Os resultados obtidos ao aplicar a abordagem de recomendação baseada em ontologias *fuzzy* de domínio são, em geral, mais precisos que os resultados da abordagem de recomendação baseada em agrupamento de dados.

$$H1W = P_{OntologiaFuzzy} > P_{Agrupamento} \Rightarrow PP_{OntologiaFuzzy} > PP_{Agrupamento}$$

$$\text{e } PC_{OntologiaFuzzy} > PC_{Agrupamento} \text{ e } PR_{OntologiaFuzzy} > PR_{Agrupamento}$$

Hipótese Alternativa (H2W) : Os resultados obtidos ao aplicar a abordagem de recomendação baseada em agrupamento de dados são, em geral, mais precisos que os resultados da abordagem de recomendação baseada em ontologias *fuzzy* de domínio.

$$H2W = P_{OntologiaFuzzy} < P_{Agrupamento} \Rightarrow PP_{OntologiaFuzzy} < PP_{Agrupamento}$$

$$\text{e } PC_{OntologiaFuzzy} < PC_{Agrupamento} \text{ e } PR_{OntologiaFuzzy} < PR_{Agrupamento}$$

3. **Seleção das Variáveis.** As variáveis **independentes** são aquelas manipuladas e controladas durante o desenvolvimento do experimento. No caso deste experimento, essas variáveis são: a abordagem de recomendação baseada em ontologias *fuzzy*; a abordagem de recomendação baseada em agrupamento de dados e os dados coletados sobre a similaridade dos dispositivos.

As variáveis **dependentes** são aquelas que estão sob análise e cujas variações devem ser observadas com base nas variáveis independentes. No

experimento realizado, a variável dependente é determinada pela precisão dos resultados obtidos.

4. **Seleção dos Participantes:** A seleção dos participantes foi feita por meio da **amostragem não probabilística por conveniência** (Wohlin et al., 2000), ou seja, a partir da seleção de indivíduos com conhecimento do domínio para participarem do experimento.
5. **Projeto do Experimento:** O Projeto do experimento descreve como os testes experimentais são organizados e executados. O estudo do experimento foi projetado de forma **aleatória** (Wohlin et al., 2000), a fim de assegurar a obtenção de resultados de diversos tipos de usuários com conhecimento do domínio.
6. **Tipo de Projeto:** O tipo do projeto do experimento foi de **um fator** (avaliação de resultados) com **dois tratamentos** (abordagem de recomendação baseada em ontologias fuzzy de domínio e abordagem de recomendação baseada em agrupamento de dados) e com **comparação pareada** (Wohlin et al., 2000). Nesse tipo de projeto, um único objeto a ser manipulado durante o estudo é utilizado em todos os tratamentos, assim como os resultados dos participantes.
7. **Instrumentação:** As medições em experimentos são conduzidas por meio dos dados coletados. Em experimentos que envolvem atividades executadas por seres humanos, os dados geralmente são coletados manualmente por meio de formulários ou entrevistas (Cirilo et al., 2011). Assim, os materiais necessários para apoiar os participantes foram previamente planejados e compreendem a definição dos objetos que seriam analisados, bem como as diretrizes para orientar os participantes no processo de avaliação.

5.2.3 Execução do Experimento

O experimento foi executado em três fases:

1. **Preparação:** Nessa fase, os materiais para a coleta dos dados definidos na instrumentação do experimento foram efetivamente elaborados. Para a coleta de dados do experimento, foi elaborado o Formulário de Avaliação de Similaridade (Apêndice A), no qual são apresentados seis conjuntos de dispositivos devidamente separados em blocos. Cada conjunto é

composto por um dispositivo inicial ou base e dez dispositivos a serem avaliados e ranqueados. Cada dispositivo é representado por meio das características mais representativas desse (Apêndice B). Assim, cada participante deveria determinar o nível de similaridade entre o dispositivo inicial e cada um dos dez dispositivos a serem avaliados e, em seguida, preencher os resultados em uma tabela de classificação no formulário.

2. **Execução:** Os participantes avaliaram a similaridade dos dispositivos considerando diferentes níveis de importância para cada característica do dispositivo. Esses níveis foram determinados por cada um dos participantes, baseado nas suas preferências. Na Tabela 5.1 mostram-se os resultados da avaliação de similaridade para o primeiro conjunto de dispositivos, feita pelos participantes na execução do experimento. A relação dos dispositivos considerados nessa avaliação encontra-se na Tabela 5.2.

Tabela 5.1: Dados Coletados da Avaliação

	Inic.01	User01	User02	User03	User04	User05	User06	User07	User08	User09	User10
Recom. 01	Disp. 08	Disp. 06	Disp. 06	Disp. 08	Disp. 10	Disp. 10	Disp. 06	Disp. 08	Disp. 02	Disp. 02	
Recom. 02	Disp. 10	Disp. 10	Disp. 10	Disp. 09	Disp. 06	Disp. 06	Disp. 02	Disp. 06	Disp. 05	Disp. 10	
Recom. 03	Disp. 02	Disp. 07	Disp. 07	Disp. 10	Disp. 09	Disp. 05	Disp. 10	Disp. 03	Disp. 03	Disp. 05	
Recom. 04	Disp. 05	Disp. 02	Disp. 05	Disp. 07	Disp. 05	Disp. 03	Disp. 09	Disp. 10	Disp. 04	Disp. 03	
Recom. 07	Disp. 04	Disp. 02	Disp. 03	Disp. 07	Disp. 02	Disp. 08	Disp. 04	Disp. 06	Disp. 06	Disp. 06	

Os dados coletados e mostrados na Tabela 5.1 estão organizados de forma a determinar qual é o ranqueamento de recomendação (Recom. 01, Recom. 02, Recom. 03, Recom. 04, Recom. 05) determinado por cada usuário relativo ao dispositivo inicial 01 (Inic. 01). Uma vez coletados os dados das avaliações feitas pelos usuários, é preciso determinar os valores de precisão correspondentes à utilização de cada uma das abordagens de recomendação em relação a esses resultados. Esses valores de precisão são determinados por meio do cálculo dos valores quantitativos associados a cada uma das métricas de avaliação de precisão nos Sistemas de Recomendação. Com o objetivo de obter os resultados da avaliação da similaridade fornecidos pelas abordagens de recomendação,

Tabela 5.2: Relação dos Dispositivos Envolvidos para o Primeiro Conjunto de Avaliação

Dispositivo	Descrição do Dispositivo
Inic. 01	BlackBerry 9520
Disp. 01	BlackBerry 0857
Disp. 02	BlackBerry 8220
Disp. 03	BlackBerry 8120
Disp. 04	BlackBerry 8110
Disp. 05	BlackBerry 8900
Disp. 06	BlackBerry 9100
Disp. 07	BlackBerry 9105
Disp. 08	BlackBerry 9300
Disp. 09	BlackBerry 9330
Disp. 10	BlackBerry 9700

foram utilizados 60 documentos que representam o conjunto de características de cada um dos dispositivos utilizados (os mesmos avaliados pelos usuários). Esses documentos foram utilizados como base para a construção de uma ontologia *fuzzy* de domínio (como apresentado no capítulo 4), a partir da qual realizou-se o procedimento de recomendação baseado em ontologia e cujos resultados constam na Tabela 5.3. Esses mesmos documentos foram utilizados como base para realizar o procedimento de recomendação baseado em agrupamento e cujos resultados também são encontrados na mesma Tabela 5.3.

Tabela 5.3: Dados Obtidos ao Avaliar os Dispositivos com as duas Abordagens de Recomendação

Inic.01	Abordagem de Recomendação baseado em Ontologias	Abordagem de Recomendação baseado em Agrupamento
Rec. 01	Disp. 08	Disp. 09
Rec. 02	Disp. 02	Disp. 04
Rec. 03	Disp. 10	Disp. 02
Rec. 04	Disp. 09	Sem Disp. No Conjunto
Rec. 05	Disp. 03	Sem Disp. No Conjunto
Rec. 06	Disp. 07	Sem Disp. No Conjunto
Rec. 07	Disp. 05	Sem Disp. No Conjunto

Já coletados os resultados, é preciso determinar os valores da precisão obtidos para cada uma das abordagens. Esses valores podem ser calculados segundo as métricas de precisão dos resultados nos Sistemas de

Recomendação, apresentadas na seção 2.4.2. Os cálculos efetuados para cada um das métricas são descritos a seguir:

Métricas de Precisão na Predição

Para o cálculo da precisão na predição da recomendação utilizou-se o valor do Erro Absoluto (MAE), no qual se tem a soma da diferença entre a avaliação do usuário e a avaliação fornecida pela abordagem de recomendação e essa é dividida pelo número de itens considerados, como ilustrado na Equação 5.1.

$$MAE = \frac{1}{|B_i|} \sum_{b_k \in B_i} |r_i(b_k) - p(b_k)| \quad (5.1)$$

Onde B_i representa o conjunto de itens avaliados pelo usuário a_i . De acordo com a Equação 5.1 e em relação às avaliações do Usuário 01 sobre o primeiro conjunto de dispositivos, as predições sobre as avaliações obtidas mostram um erro absoluto de 0.8 para a abordagem de Recomendação com ontologias e de 3.0 para a abordagem de Recomendação por Agrupamento, como ilustrado na Figura 5.1. Na Figura 5.1(a), ilustra-se a marcação efetuada pelo Usuário 01 para um dispositivo, ou seja, o conjunto de dispositivos que ele considerou mais similares com o dispositivo base. A primeira coluna mostra uma lista de 1 a 5, a qual representa os 5 dispositivos que o usuário selecionou. A segunda coluna mostra o equivalente ao dispositivo selecionado na ordem obtida pelo Sistema de Recomendação. Por exemplo, o dispositivo que o usuário marcou como o segundo mais similar, o Sistema de Recomendação marcou como o terceiro. A terceira e a quarta coluna mostram o ranqueamento dos dispositivos considerando uma quantidade de estrelas, por exemplo, para os dois primeiros dispositivos marcados pelos usuários atribui-se um valor 5, para o terceiro, um valor 4, para o seguinte, 3 e para o último, 2. Na quarta e quinta coluna observa-se o mesmo ranqueamento porém dessa vez considera-se a marcação atribuída pelo Sistema de Recomendação. Esse ranqueamento por estrelas serve para determinar o MAE de cada abordagem, segundo o *rating* de cada sistema versus o *rating* do usuário.

Os resultados do cálculo do valor do Erro Absoluto para cada uma das abordagens, em relação a cada uma das avaliações feitas pelos usuários para cada conjunto de dispositivos constam no Apêndice C.

RECOMENDAÇÃO ONTOLOGIAS					
Usuário 01	Recomendação Ontologias	Rating Usuário		Rating Ontologias	
1	1	5	☆	5	☆
2	3	5	☆	4	☆
3	2	4	☆	5	☆
4	7	3	☆	2	☆
5	6	2	☆	3	☆
6	5	Sem	☆	Sem	☆
7	4	Sem	☆	Sem	☆
MAE				0.80	

RECOMENDAÇÃO AGRUPAMENTO					
Usuário 01	Recomendação Agrupamento	Rating Usuário		Rating Agrupamento	
1	Sem Disp.	5	☆	Sem	☆
2	Sem Disp.	5	☆	Sem	☆
3	3	4	☆	4	☆
4	Sem Disp.	3	☆	Sem	☆
5	Sem Disp.	2	☆	Sem	☆
6	Sem Disp.	Sem	☆	Sem	☆
7	1	Sem	☆	Sem	☆
MAE				3.00	

Figura 5.1: Cálculo do valor de erro absoluto em relação às avaliações do usuário 01 sobre o primeiro conjunto de dispositivos

Métricas de Precisão na Classificação Para o cálculo da precisão na classificação dos itens recomendados foram utilizados os valores de precisão e de cobertura amplamente utilizados na avaliação de sistemas de recuperação de informação e também de sistemas de recomendação. A precisão é definida como a proporção dos itens relevantes recuperados dentre o conjunto de itens recuperados, como ilustrado na Equação 5.2,

$$Precisao = \frac{B_{rs}}{B_s} \quad (5.2)$$

Onde B_{rs} representa o número de itens relevantes recuperados e B_s representa o número total de itens recuperados. A cobertura é definida como a proporção de itens relevantes recuperados dentre o conjunto total de itens relevantes existentes, como ilustrado na Equação 5.3,

$$Revocacao = \frac{B_{rs}}{B_r} \quad (5.3)$$

Onde B_{rs} representa o número de itens relevantes recuperados e B_r repre-

senta o número de itens relevantes existentes no conjunto. Baseado nas Equações 5.2 e 5.3 para o primeiro conjunto de dispositivos, as predições sobre a classificação obtida para cada uma das abordagens de recomendação apresentam os seguintes valores: 0.56 de precisão média e 0.78 de cobertura média para a abordagem de recomendação com ontologias e 0.33 de precisão média e 0.46 de cobertura média para a abordagem de recomendação por agrupamento, como ilustrado na Figura 5.2. Isso demonstra maior precisão e maior cobertura para as recomendações feitas por meio da abordagem que utiliza ontologias.

	User01		User02		User03		User04		User05	
	Onto.	Agrup.	Onto.	Agrup.	Onto.	Agrup.	Onto.	Agrup.	Onto.	Agrup.
Nrs	5	2	3	3	3	3	4	2	4	2
Nrn	0	3	2	2	2	2	1	3	1	3
Nis	2	5	4	4	4	4	3	5	3	5
Nr	5		5		5		5		5	
Ns	7		7		7		7		7	

	User06		User07		User08		User09		User10	
	Onto.	Agrup.	Onto.	Agrup.	Onto.	Agrup.	Onto.	Agrup.	Onto.	Agrup.
Nrs	4	2	5	3	4	2	3	2	4	1
Nrn	1	3	0	2	1	3	2	3	1	4
Nis	3	5	2	4	3	5	4	5	3	6
Nr	5		5		5		5		5	
Ns	7		7		7		7		7	

	Ontologia		Agrupamento	
	Precisão	Cobertura	Precisão	Cobertura
User01	0.71	1.00	0.29	0.40
User02	0.43	0.60	0.43	0.60
User03	0.43	0.60	0.43	0.60
User04	0.57	0.80	0.29	0.40
User05	0.57	0.80	0.29	0.40
User06	0.57	0.80	0.29	0.40
User07	0.57	0.80	0.29	0.40
User08	0.71	1.00	0.43	0.60
User09	0.57	0.80	0.29	0.40
User10	0.43	0.60	0.29	0.40
Média	0.56	0.78	0.33	0.46

Figura 5.2: Cálculo dos valores de precisão e cobertura média em relação ao primeiro conjunto de dispositivos

Os resultados do cálculo da precisão e da cobertura média para cada uma das abordagens, em relação a cada uma das avaliações feitas pelos usuários para cada conjunto de dispositivos, encontram-se no Apêndice D. Com o propósito de obter uma melhor visualização sobre o comportamento dos valores que representam a precisão, a cobertura e a *F-measure* para cada abordagem, foram feitas várias análises individuais sobre cada uma das avaliações fornecidas pelos usuários para cada um dos conjuntos de dispositivos. Nessas análises, foram estabelecidos alguns pontos de corte (*cut-offs*) na sequência dos resultados recomendados, os quais permitiram determinar a evolução em termos de precisão e de cobertura de cada uma das abordagens. Na Figura 5.3 é ilustrado o comportamento das abordagens em relação às avaliações fornecidas pelo primeiro usuário, para o primeiro conjunto de dispositivos.

O cálculo e a representação gráfica da precisão, da cobertura e da *F-measure* para cada um das abordagens em relação a cada uma das ava-

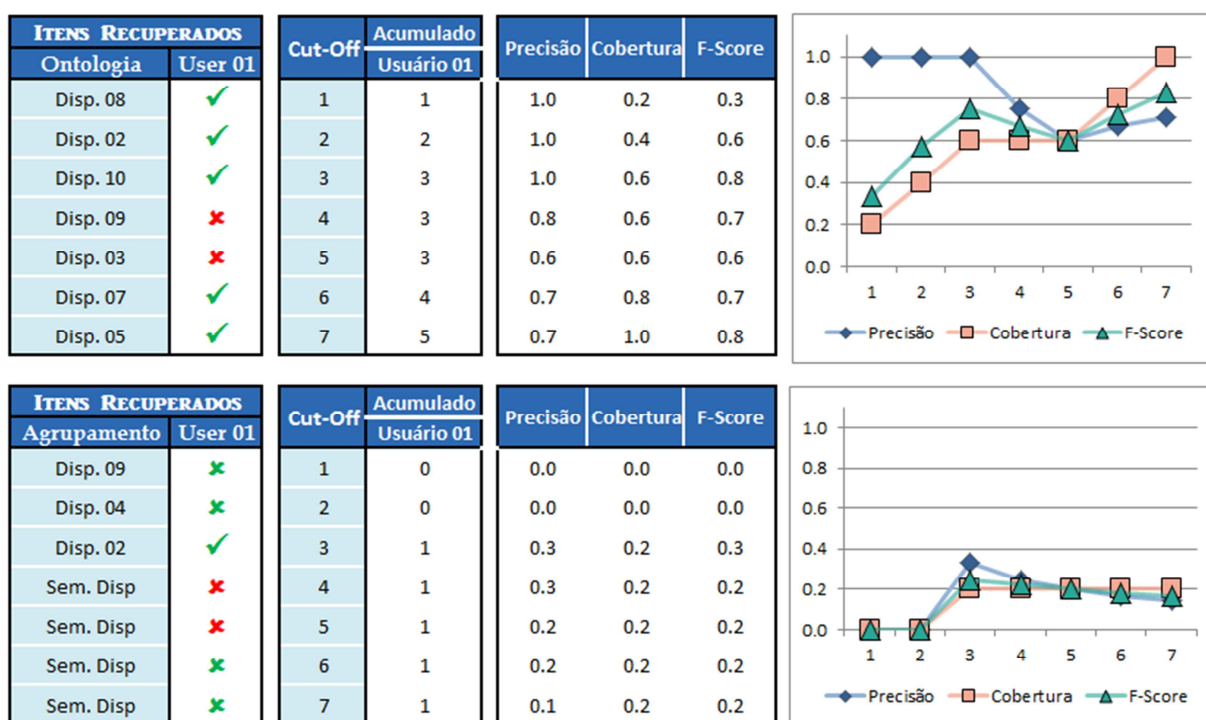


Figura 5.3: Cálculo e representação gráfica dos valores de precisão, cobertura e F-measure para cada uma das abordagens, em relação à avaliação do usuário 01 sobre o primeiro conjunto de dispositivos

liações feitas pelos usuários para cada conjunto de dispositivos são apresentados no Apêndice E.

Métricas de Precisão no Ranqueamento Para o cálculo da precisão no ranqueamento dos itens recomendados, utilizou-se como métrica de avaliação, o valor do coeficiente de *Spearman* (ρ), também conhecido como *Spearman's rho*. Esse coeficiente é definido como uma medida não paramétrica de dependência estatística entre duas variáveis (Herlocker et al., 2000). O cálculo da medida do coeficiente é dado pela Equação 5.4.

$$\rho = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)} \quad (5.4)$$

Onde d_i representa a diferença entre os valores de cada avaliação de ranqueamento e n representa o número total de itens ranqueados. De acordo com essa Equação e levando em conta a avaliação do primeiro conjunto de dispositivos realizada pelo Usuário 01, as predições obtidas por cada uma das duas abordagens de recomendação para o ranqueamento dos

itens apresentam os seguintes valores: um coeficiente (ρ) de 0.607 para a abordagem de recomendação com ontologias e um coeficiente (ρ) de -1.107 para a abordagem de recomendação por agrupamento, como ilustrado na Figura 5.4. Como se pode observar, o valor da abordagem de recomendação com ontologias é mais próximo de 1 e, portanto, reflete melhor as preferências do usuário.

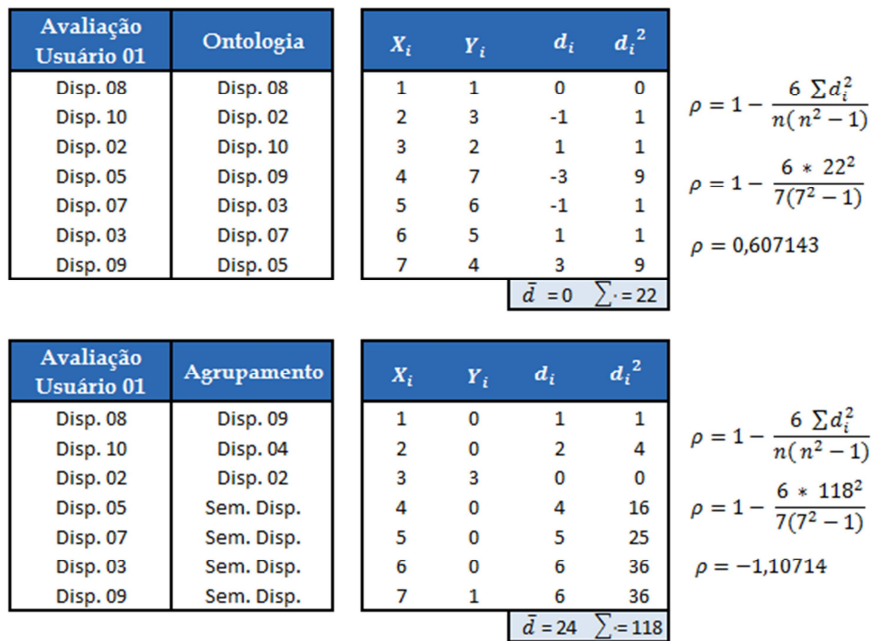


Figura 5.4: Cálculo dos valores de Spearman para cada uma das abordagens, em relação à avaliação do usuário 01 sobre o primeiro conjunto de dispositivos

- Validação:** Após a execução do experimento, verificou-se se os dados registrados pelos usuários eram razoáveis, bem como se foram inseridos da forma correta. Durante a análise, observou-se que os resultados obtidos ao realizar os cálculos da precisão no terceiro conjunto de dispositivos eram inconsistentes, devido ao fato de não existir interseção de dados entre os resultados obtidos ao aplicar as abordagens de recomendação exploradas nesse experimento e os resultados avaliados pelos usuários. Sem levar em conta a diferença gerada por esses dados, os tratamentos foram aplicados corretamente, conforme planejado no projeto experimental.

5.3 Análise e Interpretação dos Resultados

A partir de uma análise preliminar dos valores de precisão descritos na seção anterior, alguns aspectos interessantes foram observados entre os resultados obtidos ao aplicar a abordagem de recuperação baseada em ontologias fuzzy e a baseada em agrupamento de dados.

Em relação ao cálculo dos valores do erro absoluto (MAE), na Figura 5.1 é ilustrada uma distribuição média desses valores para cada uma das abordagens, em relação a cada um dos conjuntos de dispositivos avaliados. Nessa representação, nota-se que os resultados obtidos pela abordagem de recomendação baseada em ontologias *fuzzy* possuem um erro menor de precisão, a partir do qual pode-se determinar que a predição da abordagem de recomendação baseada em ontologias *fuzzy* tem uma maior correlação com os dados avaliados pelos usuários.

Em relação ao cálculo dos valores de precisão e de cobertura das abordagens de recomendação, na Figura 5.2 é ilustrado o comportamento de cada uma delas em relação à avaliação feita pelo primeiro usuário para o primeiro conjunto de dispositivos. Essa representação simboliza a tendência das outras avaliações.

Observa-se que os valores finais da precisão e da cobertura obtidos na abordagem de recomendação utilizando ontologias *fuzzy* são superiores aos obtidos na abordagem que utiliza agrupamento de dados. Esses valores indicam que a abordagem de recomendação que utiliza ontologias *fuzzy* recuperou, para o usuário, uma maior quantidade de resultados relevantes com um menor número de resultados irrelevantes de interferência.

Em relação ao cálculo dos valores de precisão no ranqueamento de cada uma das abordagens de recomendação, na Figura 5.4 é ilustrada uma distribuição média dos valores dos coeficientes de *Spearman* obtidos na avaliação de cada conjunto de dispositivos. Pode-se observar, na Figura 5.4, que os coeficientes de *Spearman* obtidos da correlação entre os resultados da abordagem de recomendação baseada em ontologias *fuzzy* e os resultados da avaliação dos usuários são menores que os coeficientes obtidos da correlação entre os resultados da abordagem de recomendação baseada em agrupamento e os resultados da avaliação dos usuários. Isso indica que o ranqueamento feito pela abordagem de recomendação baseada em ontologias *fuzzy* mostra uma maior correlação com o ranqueamento estabelecido pelo usuário.

Após essas observações iniciais, passou-se para a obtenção das conclusões

do estudo, a qual envolve a caracterização dos dados por meio do teste das hipóteses. Os métodos estatísticos para o teste das hipóteses dependem da qualidade dos dados de entrada para a obtenção de bons resultados. Se os dados sobre os quais um método estatístico é aplicado estiverem incorretos, os resultados do método também estarão. É necessário, portanto, remover das amostras todos os dados redundantes que podem levar a resultados errôneos. Falhas no conjunto de dados geralmente ocorrem por conta de erros sistemáticos (e.g. erros durante a transcrição dos dados) ou pela presença de *outliers*, comparados a “pontos fora da curva”, pontos muito menores ou muito maiores do que o esperado em relação aos demais dados do conjunto. Conforme descrito na análise dos resultados de precisão, um *outlier* foi encontrado nos cálculos, o que gerou a necessidade da sua remoção para a obtenção de bons resultados estatísticos. É importante notar que esses dados foram identificados na etapa de “Validação dos Dados” da fase de “Execução do estudo”, cujo objetivo era identificar dados inválidos com base na execução do experimento.

5.3.1 Teste das Hipóteses

O objetivo desse passo é verificar, com algum grau de significância, se é possível rejeitar a hipótese nula (H_0) em favor de alguma das hipóteses alternativas (H_1 ou H_2), com base no conjunto de dados obtido.

Inicialmente, os dados obtidos no experimento foram analisados e, aparentemente, a precisão dos resultados que utilizaram a abordagem baseada em ontologias *fuzzy* foi maior que a precisão dos resultados que utilizaram a abordagem baseada em agrupamento. A fim de comprovar esse efeito de forma estatística, aplicou-se o *Paired T-test* (Montgomery, 1999). Esse teste paramétrico é usado para comparar dois conjuntos de dados utilizando o mesmo conjunto de medição e analisar se diferem um do outro e, assim, provar que o efeito hipotético foi demonstrado. No caso do estudo realizado, as amostras foram constituídas pelos dados referentes às medidas de precisão obtidas tanto na utilização da abordagem com ontologias *fuzzy*, quanto na abordagem por agrupamento.

Na Equação 5.5 é mostrada a fórmula do *Paired t-test*. Trata-se de uma razão: no numerador tem-se a diferença das médias e no denominador tem-se a derivação *standard* das diferenças e do número de itens dados do conjunto (n), conforme o cálculo mostrado na Equação 5.6

$$t_0 = \frac{d}{s_d \sqrt{n}} \quad (5.5)$$

$$S_d = \sqrt{\frac{\sum_{i=1}^n (d_i - d)^2}{n - 1}} \quad (5.6)$$

Após o cálculo de t_0 , deve-se verificar a tabela padrão de distribuição de probabilidade estatística *t de Student*, mostrada no Apêndice F (Gosset, 1947) e analisar se a razão calculada é suficientemente grande para rejeitar a hipótese de que a diferença existente entre os conjuntos seja mera causalidade. Define-se, portanto, um grau de significância α , que representa o “nível de risco”. Assim pode-se afirmar que há diferença entre as amostras e rejeitar a hipótese nula, por exemplo, se considerado um valor $\alpha = 0.05$, isso significa que se assume um risco de 5% de encontrar uma diferença significativa entre as médias dos conjuntos, mesmo que essa diferença seja um Falso Positivo. O nível de confiança do resultado do teste, neste caso, é de 95%. Também é necessário estabelecer o grau de liberdade (gl) para o teste, o qual corresponde à quantidade de dados nos conjuntos menos um ($n - 1$).

Assim, tendo-se $t_{0,\alpha}$ e gl , observa-se o valor de t na tabela padrão para determinar se o valor de t_0 é alto o suficiente para ser significativo. O valor t da tabela reflete a probabilidade de ter sido obtida a diferença observada entre as médias dos conjuntos numa situação onde a precisão dos resultados de ambas as abordagens se equiparam, i.e., a hipótese nula é verdadeira. Se essa probabilidade for muito baixa, pode-se concluir que o resultado observado no estudo é estatisticamente relevante. Nesse caso, se o valor t da tabela, em relação ao nível de risco escolhido e ao grau de liberdade for inferior ao t_0 calculado, a hipótese nula é rejeitada. Caso contrário, a hipótese nula não é rejeitada e nenhuma conclusão pode ser obtida a partir do experimento.

O efeito na variável dependente do experimento (precisão dos resultados) envolveu os três aspectos seguintes: precisão na predição dos resultados (representada pelo valor do erro absoluto - MAE), precisão da classificação dos resultados (representada pela medida *F-measure* entre a precisão e a cobertura) e precisão do ranqueamento dos resultados (representado pelo coeficiente de *Spearman*). Assim, a aplicação do *Paired T-test* ao conjunto amostral de dados foi realizada em três etapas. Na primeira, fez-se a comparação dos conjuntos de dados referentes aos valores de precisão na predição dos resultados de cada abordagem. Na segunda, fez-se a comparação entre os conjuntos de dados referentes aos valores de precisão na classificação dos resultados de

cada abordagem. Já na terceira, a comparação foi feita entre os conjuntos de dados referentes aos valores de precisão no ranqueamento dos resultados de cada abordagem. O teste da hipótese nula baseou-se na combinação dos critérios de rejeição das três etapas do teste, de modo que H_0 apenas seria rejeitada se, e somente se, pudesse ser rejeitada nas três etapas do teste. Para os propósitos deste estudo, em cada etapa do teste utilizou-se o menor grau de significância α com o qual fosse possível rejeitar a hipótese nula, assumindo-se um grau de significância máximo de 5%. Nas Tabelas 5.4, 5.5 e 5.6 mostra-se a organização do teste realizado.

Tabela 5.4: Organização da Primeira etapa do *Paired t-test* para a precisão na predição (ρ)

	Precisão na predição (ρ)
Entrada	Duas amostras pareadas: <ul style="list-style-type: none"> • $X_{\rho_{Ontologia}} = 0.80, 1.60, 2.20, 1.20, 1.60, 2.80, 1.00, \dots, 1.60$ • $Y_{\rho_{Agrupamento}} = 3.00, 3.80, 2.80, 3.20, 4.20, 1.80, 3.20, \dots, 1.40$
$H_{0\rho}$	$H_{0\rho} : \mu_{d\rho} = 0$ onde $d_{i\rho} = X_{i\rho_{Ontologia}} - Y_{i\rho_{Agrupamento}}$, i.e., o valor médio da diferença dos valores das abordagens é zero.
Cálculo de $t_{0\rho}$	$t_{0\rho} = \frac{\bar{d}}{\frac{S_{d\rho}}{\sqrt{n}}}$, onde: $S_{d\rho} = \sqrt{\frac{\sum_{i=1}^n d_{i\rho}^2 - \bar{d}^2}{n-1}}$
Critério de rejeição de $H_{0\rho}$	<ul style="list-style-type: none"> • $H_{1\rho} : \mu_{\rho_{Ontologia}} < \mu_{\rho_{Agrupamento}}$; • $H_{2\rho} : \mu_{\rho_{Ontologia}} > \mu_{\rho_{Agrupamento}}$; • Logo, $\mu_{d\rho} \neq 0$: rejeitar $H_{0\rho}$ se $t_{0\rho} > t_{\alpha/2, gl\rho}$

1. **Primeira Etapa:** Conforme o conjunto de dados de precisão na predição, tem-se 60 amostras pareadas presentes nos conjuntos $|X_{\rho_{Ontologia}}|$ e $|Y_{\rho_{Agrupamento}}|$ e o valor médio da diferença entre elas é $\mu_{d\rho} = 1.18$. Pode-se, então, obter $|S_{d\rho}| = 0.71$ e $t_{0\rho} = 12.72$.

O número de graus de liberdade é $gl\rho = |X_{\rho_{Ontologia}}| - 1 = 60 - 1 = 59$ e utilizou-se $\alpha = 0.05$. Na tabela padrão de distribuição de probabilidade t de *Student*, verifica-se que $t_{0.025, 59} = 2.390$. Desde que $|t_{0\rho}| > t_{0.025, 59}$ é possível rejeitar a hipótese nula $H_{0\rho}$ com 5% de significância.

2. **Segunda Etapa:** Conforme o conjunto de dados de precisão na classificação, tem-se 60 amostras pareadas presentes nos conjuntos $|X_{\vartheta_{Ontologia}}|$ e $|Y_{\vartheta_{Agrupamento}}|$ e o valor médio da diferença entre elas é $\mu_{d\vartheta} = 0.34$. Pode-se,

Tabela 5.5: Organização da Segunda etapa do *Paired t-test* para a precisão na classificação (ϑ)

	Precisão na classificação (ϑ)
Entrada	Duas amostras pareadas: <ul style="list-style-type: none"> • $X_{\vartheta_{Ontologia}} = 0.72, 0.36, 0.54, 0.36, 0.90, 0.54, 0.72, \dots, 0.36$ • $Y_{\vartheta_{Agrupamento}} = 0.18, 0.36, 0.36, 0.54, 0.72, 0.90, 0.18, \dots, 0.72.$
$H_{0\vartheta}$	$H_{0\vartheta} : \mu_{d\vartheta} = 0$ onde $d_{i\vartheta} = X_{i\vartheta_{Ontologia}} - Y_{i\vartheta_{Agrupamento}}$, i.e., o valor médio da diferença dos valores das abordagens é zero.
Cálculo de $t_{0\vartheta}$	$t_{0\vartheta} = \frac{\bar{d}}{\frac{S_{d\vartheta}}{\sqrt{n}}}$, onde: $S_{d\vartheta} = \sqrt{\frac{\sum_{i=1}^n d_{i\vartheta}^2 - \bar{d}^2}{n-1}}$
Critério de rejeição de $H_{0\vartheta}$	<ul style="list-style-type: none"> • $H_{1\vartheta} : \mu_{\vartheta_{Ontologia}} < \mu_{\vartheta_{Agrupamento}}$; • $H_{2\vartheta} : \mu_{\vartheta_{Ontologia}} > \mu_{\vartheta_{Agrupamento}}$; • Logo, $\mu_{d\vartheta} \neq 0$: rejeitar $H_{0\vartheta}$ se $t_{0\vartheta} > t_{\alpha/2, gl\vartheta}$

Tabela 5.6: Organização da Terceira etapa do *Paired t-test* para a precisão no ranqueamento (π)

	Precisão no Ranqueamento (π)
Entrada	Duas amostras pareadas: <ul style="list-style-type: none"> • $X_{\pi_{Ontologia}} = 0.75, 0.42, 0.35, 0.36, 0.46, -0.23, 0.10, \dots, 0.71$ • $Y_{\pi_{Agrupamento}} = -0.07, 0.23, -0.73, 0.39, -0.33, 0.50, 0.02, \dots, -0.46.$
$H_{0\pi}$	$H_{0\pi} : \mu_{d\pi} = 0$ onde $d_{i\pi} = X_{i\pi_{Ontologia}} - Y_{i\pi_{Agrupamento}}$, i.e., o valor médio da diferença dos valores das abordagens é zero.
Cálculo de $t_{0\pi}$	$t_{0\pi} = \frac{\bar{d}}{\frac{S_{d\pi}}{\sqrt{n}}}$, onde: $S_{d\pi} = \sqrt{\frac{\sum_{i=1}^n d_{i\pi}^2 - \bar{d}^2}{n-1}}$
Critério de rejeição de $H_{0\pi}$	<ul style="list-style-type: none"> • $H_{1\pi} : \mu_{\pi_{Ontologia}} < \mu_{\pi_{Agrupamento}}$; • $H_{2\pi} : \mu_{\pi_{Ontologia}} > \mu_{\pi_{Agrupamento}}$; • Logo, $\mu_{d\pi} \neq 0$: rejeitar $H_{0\pi}$ se $t_{0\pi} > t_{\alpha/2, gl\pi}$

então, obter $|S_{d\vartheta}| = 0.19$ e $t_{0\vartheta} = 13.76$.

O número de graus de liberdade é $gl\vartheta = |X_{\vartheta_{Ontologia}}| - 1 = 60 - 1 = 59$ e utilizou-se $\alpha = 0.05$. Na tabela padrão de distribuição de probabilidade *t* de *Student*, verifica-se que $t_{0.025, 59} = 2.390$. Desde que $|t_{0\vartheta}| > t_{0.025, 59}$ é possível rejeitar a hipótese nula $H_{0\vartheta}$ com 5% de significância.

3. Terceira Etapa: Conforme o conjunto de dados de precisão no ranquea-

mento, tem-se 60 amostras pareadas presentes nos conjuntos $|X_{\pi_{Ontologia}}|$ e $|Y_{\pi_{Agrupamento}}|$ e o valor médio da diferença entre elas é $\mu_{d_{\pi}} = 0.48$. Pode-se, então, obter $|S_{d_{\pi}}| = 0.35$ e $t_{0_{\pi}} = 10.53$.

O número de graus de liberdade é $gl_{\pi} = |X_{\pi_{Ontologia}}| - 1 = 60 - 1 = 59$ e utilizou-se $\alpha = 0.05$. Na tabela padrão de distribuição de probabilidade *t* de *Student*, verifica-se que $t_{0.025,59} = 2.390$. Desde que $|t_{0_{\pi}}| > t_{0.025,59}$ é possível rejeitar a hipótese nula $H_{0_{\pi}}$ com 5% de significância.

5.3.2 Conclusões do Experimento

Uma vez rejeitada a hipótese nula (H_{W0}), pode-se afirmar que as diferenças observadas na precisão dos resultados obtidos utilizando cada uma das abordagens de recomendação possuem uma significância estatística.

Conforme ilustrado na Figura 5.1, o valor de precisão na predição dos resultados obtidos a partir da utilização da abordagem baseada em ontologias *fuzzy* foi maior que o valor de precisão na predição dos resultados a partir da abordagem baseada em agrupamento. Pode-se evidenciar também o comportamento para os valores de precisão na classificação e no ranqueamento dos resultados, Figura 5.2 e Figura 5.4, respectivamente. Pode-se observar que a abordagem de recomendação baseada em ontologias *fuzzy* apresenta um valor maior de precisão. Esses dados evidenciam que a hipótese alternativa H_{1W} pode ser validada em detrimento da hipótese H_{2W} , ou seja, há indícios para afirmar que os resultados obtidos pela abordagem baseada em ontologias *fuzzy* são, em geral, mais precisos que os resultados obtidos pela abordagem baseada em agrupamento, e não ao contrário.

Finalmente, considerando que o experimento foi realizado *in-vitro*, sob condições controladas, é importante enfatizar que as conclusões a respeito dos resultados observados neste trabalho se restringem ao escopo de recomendação de dispositivos móveis, no contexto da sua similaridade e baseada nas suas características expressas no WURFL¹. A fim de poder generalizar o fenômeno observado em um contexto mais amplo, é necessário que novos estudos sejam realizados em ambientes *in-vivo* com um domínio maior, comparando-se também o uso da abordagem baseada em ontologias *fuzzy* em relação a outras abordagens que utilizem outro tipo de metodologia para a recomendação. Isso permitirá obter uma validação mais abrangente das hipóteses de pesquisa.

¹<http://wurfl.sourceforge.net/>

5.3.3 Ameaças à Validade

Após determinada a validade dos resultados, uma das questões fundamentais de um experimento é determinar o quão válidos eles são. Devido a isso, é importante considerar a questão da validade da avaliação desde a fase de planejamento, de forma a guiar a execução do experimento com o objetivo de obter uma validade adequada dos resultados (Wohlin et al., 2000).

As principais ameaças que podem colocar em risco a validade de um experimento se classificam em quatro tipos: de conclusão, interna, de construção e externa (Wohlin et al., 2000). Essas ameaças foram tratadas no experimento realizado, conforme segue:

Validade de Conclusão : Refere-se às questões que afetam a habilidade de tirar conclusões a respeito do tratamento nas variáveis dependentes, por exemplo, a escolha do método estatístico adequado para a análise, análise, os cuidados tomados na execução e na medição do experimento. No experimento, adotou-se o *Paired T-test* como método estatístico para a análise. Essa decisão foi tomada por se tratar de um experimento do tipo **um fator** (avaliação de resultados) com **dois tratamentos** (abordagem de recomendação baseado em ontologias *fuzzy* de domínio e abordagem de recomendação baseada em agrupamento de dados) com **comparação pareada**. Esse teste é o mais adequado para projetos desse tipo, nos quais o objetivo é comparar as médias dos resultados de dois tratamentos distintos obtidas a partir mesmo conjunto de dados. Particularmente, por ser um teste paramétrico, sua validade é, em geral, maior do que a de testes não-paramétricos, o que também justifica o seu uso para obter resultados mais precisos no estudo. Além disso, as medidas utilizadas no experimento envolveram dados diretos (precisão), que independem de julgamento humano. Isso aumenta a confiabilidade dos dados coletados no experimento.

Validade Interna : Refere-se às questões que afetam a habilidade de assegurar que os resultados foram obtidos, de fato, em decorrência dos tratamentos e não por coincidência ou pelo efeito de outro fator que não foi medido ou não se pôde controlar, por exemplo, modo como os participantes são selecionados ou o modo de apresentação dos dados para os participantes avaliarem. Nessa pesquisa, procurou-se realizar o experimento com usuários de dispositivos móveis com conhecimento geral das suas características. Assim, assumiu-se que são representativos para a

população de potenciais usuários de um Sistema de Recomendação de dispositivos móveis. Além disso, no formulário de avaliação entregue a cada um dos participantes não foi considerado a nome do modelo nem a imagem do dispositivo, isso para evitar que a parte visual interferisse nas avaliações dos participantes. Além disso, os participantes realizaram a avaliação de forma individual, evitando, assim, que pudessem alterar os valores da avaliação para que fossem coerentes com os valores dos outros participantes, o que certamente afetaria a precisão dos resultados.

Validade de Construção : Refere-se às questões que afetam a habilidade de generalizar o resultado do experimento ao conceito ou à teoria envolvida no estudo, por exemplo, a definição adequada das medidas e dos tratamentos. No estudo realizado, o objetivo foi comparar duas abordagens de recomendação considerando o seu impacto na precisão dos resultados. Para os propósitos do experimento, a abordagem mais precisa é aquela que apresenta valores de precisão de predição, de classificação e de ranqueamento mais altos em relação à outra. Portanto, os dados para a coleta e para os tratamentos foram definidos de modo que fosse possível realizar adequadamente a análise da precisão das abordagens em conformidade com os objetivos do estudo.

Validade Externa : Refere-se às questões que afetam a habilidade de generalizar os resultados do experimento para um contexto mais amplo que o selecionado para o estudo. Nesse caso, existem três riscos principais: a escolha errônea dos participantes, a condução do experimento em um ambiente inapropriado, e o desenrolar do estudo em um período de tempo que afete os resultados. No caso, os participantes podem ser considerados representativos de uma população de possíveis usuários de Sistemas de Recomendação com interesse em dispositivos móveis. O experimento foi realizado por meio de avaliações, conduzidas pelos participantes de forma individual e por via eletrônica, na qual não era necessário permanecer em uma ubiquação específica durante o experimento. Com relação às questões temporais, o experimento foi planejado de modo que a avaliação feita pelos participantes não tivesse um tempo limite de entrega. Isso permitiu que eles realizassem uma análise precisa das características dos dispositivos com o intuito de fornecer uma avaliação com um alto grau de confiança.

5.4 Considerações Finais

Este capítulo apresentou a avaliação da abordagem proposta. Essa seguiu a metodologia experimental para realizar um estudo comparativo entre uma abordagem de recomendação baseada em ontologias *fuzzy* e uma abordagem de recomendação baseada em agrupamento de dados.

Os resultados favoráveis da avaliação do trabalho corroboram com as expectativas iniciais do processo desenvolvido, no qual busca-se o aumento da precisão dos três parâmetros avaliados no experimento: precisão na predição, na classificação e no ranqueamento.

Conclusão

Neste trabalho descrevem-se os principais fundamentos dos Sistemas de Recomendação, dentre eles o processo de recomendação, as suas principais técnicas de recomendação e as métricas de avaliação. Além disso, encontra-se uma descrição do estado da arte de técnicas que abordam a construção de ontologias *fuzzy* e de Sistemas de Recomendação que usam ontologias de usuário e permitem a propagação do interesse. Foram também descritos os métodos e as técnicas utilizadas para a elaboração de uma abordagem de recomendação que usa como base uma ontologia *fuzzy* de domínio e uma ontologia *crisp* de preferências de usuário. Essa abordagem é baseada em duas etapas: (1) Engenharia da Ontologia, que propôs uma série de fluxos que permitem fuzzificar uma ontologia *crisp* existente e uma estratégia de representação das preferências do usuário; e (2) Engenharia do Sistema de Recomendação, na qual foram apresentadas as atividades necessárias para a implementação de uma sistema de recomendação que permita utilizar as ontologias produzidas na engenharia de ontologia.

Apresentou-se também, um estudo de caso, no qual são detalhadas as etapas da Engenharia da ontologia e da Engenharia do Sistema de Recomendação aplicadas ao domínio de *e-commerce*, especificamente dos dispositivos móveis.

A fim de avaliar a abordagem de recomendação que usa uma ontologia *fuzzy* de domínio e uma ontologia *crisp* de preferências de usuário, essa foi comparada a uma abordagem que utiliza agrupamento de dados. Para realizar a avaliação, foram definidas três métricas: (1) precisão na predição, que

mede o erro absoluto existente entre os dados recuperados e os dados considerados pelo usuário, (2) precisão na classificação, que mede a quantidade de resultados relevantes recuperados, e (3) precisão no ranqueamento, que mede a correlação existente entre o ranking dos sistemas de recomendação e o ranking estabelecido pelo usuário.

Os resultados indicam que o uso da abordagem que utiliza uma ontologia *fuzzy* de domínio e uma ontologia *crisp* de preferências de usuário contribui significativamente para a melhoria dos resultados das recomendações. Em particular, a aplicação da abordagem baseada nas ontologias melhora a precisão nas três métricas definidas.

6.1 Principais Resultados e Contribuições

Uma contribuição deste trabalho consistiu na adaptação da metodologia UPON, aqui denominada UPFON, com o intuito de permitir a fuzzificação de uma ontologia *crisp* existente. Foram definidos uma série de passos a fim de estabelecer a similaridade entre os conceitos da ontologia. Outra contribuição consistiu na especificação de um método que considera a propagação dos interesses do usuário nos processos de recomendação. Esse método juntamente com a ontologia *fuzzy* de domínio permitem desenvolver a filtragem dos Sistemas de Recomendação.

A partir dos experimentos realizados conclui-se que as recomendações efetuadas com o uso das ontologias apresentaram um melhor desempenho. Embora os experimentos tenham sido realizados em um único domínio, outros podem ser utilizados, desde que seja possível usar um conjunto específico de documentos para a construção da ontologia *fuzzy*.

Algumas das contribuições deste trabalho tiveram seus resultados publicados em eventos científicos da área, como apresentado a seguir:

- Urquizo, A., Santos, M. T. P., e Prado, A. F. (2012). Upfon: Unified process for building fuzzy ontology. International Conference on Fuzzy Systems and Knowledge Discovery. Onde foi apresentada a adaptação UPFON (Urquizo et al., 2012).
- Souza, D. F., Urquizo, A., Lopez, W., e Prado, A. F. (2011). Amrp: Ambiente para o monitoramento remoto de pacientes. Centro Latinoamericano de Estudios en Informática. Onde foram apresentados os estudos iniciais sobre dispositivos móveis (Souza et al., 2011).

A fim de complementar o presente trabalho, as principais limitações e os trabalhos futuros são sugeridos, nas seguintes seções.

6.2 Principais Limitações

A partir de uma análise crítica do trabalho desenvolvido, foram identificadas algumas limitações. A abordagem de recomendação proposta visa apoiar o desenvolvimento de aplicações de recomendação mais precisas. Vale ressaltar que o escopo dessa abordagem restringiu-se ao tipo de filtragem baseada em conteúdo que fazem as recomendações de acordo com os relacionamentos existentes entre os conceitos do domínio.

Outra limitação diz respeito à adaptação UPFON para a fuzzificação de ontologias, a qual necessita de um conjunto de documentos estruturados e específicos de domínio, como parâmetro de entrada à adaptação UPFON. Essa estrutura corresponde a uma representação atributo-valor para cada um dos conceitos presentes na ontologia a ser fuzzificada. Uma alternativa seria tornar mais flexíveis os formatos desses documentos de entrada.

A última limitação deste trabalho refere-se ao estudo experimental apresentado no capítulo de avaliação. Os experimentos foram realizados com um número reduzido de usuários no domínio dos dispositivos móveis. Para melhorar a confiabilidade dos resultados é necessário realizar os experimentos com um maior número de usuários.

6.3 Trabalhos Futuros

Ao longo do desenvolvimento deste trabalho, algumas oportunidades de melhoria, tanto do processo quanto do suporte computacional, foram destacadas. Além disso, novas oportunidades de pesquisa foram identificadas. Alguns possíveis trabalhos futuros envolvem:

A extensão da abordagem ObRA com outros tipos de técnicas de filtragem, como a filtragem colaborativa, que reutiliza a ontologia crisp de preferências de usuário gerada a partir da ontologia de domínio com o intuito de associar um conjunto de preferências a um determinado grupo de usuários que compartilham interesses.

Com relação à fuzzificação de ontologias, técnicas mais avançadas de mineração de dados poderiam ser exploradas na etapa de pré-processamento dos

documentos, evitando, assim, a necessidade de coletar documentos específicos de domínio em um formato estruturado.

Por fim, a abordagem pode ser aplicada a fim de atender outros domínios e não somente o dos dispositivos móveis, foco deste trabalho.

Referências Bibliográficas

Adomavicius, G., Sankaranarayanan, R., Sen, S., e Tuzhilin, A. (2005). Incorporating contextual information in recommender systems using a multi-dimensional approach. *ACM Transactions on Information Systems*, 23:103–145. Não citado no texto.

Adomavicius, G. e Tuzhilin, A. (2005a). Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. *IEEE Transactions on Knowledge and Data Engineering*, 17:734–749. Não citado no texto.

Adomavicius, G. e Tuzhilin, A. (2005b). Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. *IEEE Transactions on Knowledge and Data Engineering*, 17:734–749. Não citado no texto.

Adonie, R., Russo, J., e Dean, R. (2005). Crossing the rubicon for an intelligent advisor. *Workshop Beyond Personalization*, 1:7–12. Não citado no texto.

Almeida, M. B. e Bax, M. P. (2003). Uma visão geral sobre ontologias: pesquisa sobre definições, tipos, aplicações, métodos de avaliação e de construção. *Ciência da Informação*, 3:7–20. Não citado no texto.

Ansari, A., Essegai, S., e Kohli, R. (2000). Internet recommendation systems. *Journal of Marketing Research*, 37:363–375. Não citado no texto.

Antoniou, G. e Van Harmelen, F. (2003). Web ontology language: Owl. In *Handbook on Ontologies in Information Systems*, páginas 67–92. Springer. Não citado no texto.

- Baeza-Yates, R. A. e Ribeiro-Neto, B. (1999). *Modern Information Retrieval*. Addison-Wesley Longman Publishing Co., Inc. Não citado no texto.
- Balabanovic, M. e Shoham, Y. (1997). Fab: Content-based, collaborative recommendation. *Communications of the ACM*, 1:66–72. Não citado no texto.
- Batista, G. E. d. A. P. A. (2003). *Pré-processamento de Dados em Aprendizado de Máquina Supervisionado*. phdtese de doutorado, ICMC-USP, São Carlos - SP. Não citado no texto.
- Belkin, N. e Croft, W. B. (1992). Information filtering and information retrieval: Two sides of the same coin? *Communications of the ACM*, 35:29–39. Não citado no texto.
- Bernaras, A., Laresgoiti, I., e Correa, J. (1996). Building and reusing ontologies for electrical network applications. In *European conference on artificial intelligence*, páginas 298–302. Não citado no texto.
- Berners-Lee, T., Hendler, J., e Lassila, O. (2001). The semantic web. *Scientific American*, 284:34–43. Não citado no texto.
- Bhogal, J., Macfarlane, A., e Smith, P. (2007). A review of ontology based query expansion. *Information Processing and Management*, 43:866–886. Não citado no texto.
- Burke, R. (2007). Hybrid web recommender systems. *The Adaptive Web*, 1:377–408. Não citado no texto.
- Cazella, S. C. (2006). *Aplicando a Relevância da Opinião de Usuários em Sistemas de Recomendação para Pesquisadores*. Tese de Doutorado, Universidade Federal do Rio Grande do Sul. Não citado no texto.
- Cena, F., Likavec, S., e Osborne, F. (2011). Propagating user interests in ontology-based user model. In *Proceedings of the 12th international conference on Artificial intelligence around man and beyond*, páginas 299–311. Springer-Verlag. Não citado no texto.
- Cerri, M. J.; Yaguinuma, C. A. (2010). The ufocore: Exploring fuzzy relations according to specific contexts. In *Internacional Conference on Software Engineering & Knowledge Engineering (SEKE)*, páginas 529–534. Não citado no texto.

- Chandrasekaram, B. e Josephson, J. R. (1997). The ontology of tasks and methods. In *Symposium on Ontological Engineering*. Não citado no texto.
- Chen, A. Y. e Mccleod, D. (2005). Collaborative filtering for information recommendation systems. In *Encyclopedia of Data Warehousing and Mining*. Idea Group. Não citado no texto.
- Cirilo, C. E., Prado, A. F., Souza, W. L., e Martinez Zaina, L. A. (2011). Experimentação do processo model driven rich ubi no desenvolvimento de interfaces ricas adaptativas. *XXV Simpósio Brasileiro de Engenharia de Software (SBES)*, 1:182–191. Não citado no texto.
- Conrado, M. S., Marcacini, R. M., Moura, M. F., e Rezende, S. O. (2009). O efeito do uso de diferentes formas de geração de termos na compreensibilidade e representatividade dos termos em coleções textuais na língua portuguesa. In *II Workshop on Web and Text Intelligence (WTI)*, páginas 1–10, São Carlos, SP-Brazil. Não citado no texto.
- Corrêa, I. e Cazella, S. (2007). Um modelo para recomendação de itens baseado em filtragem colaborativa para dispositivos móveis. In *Workshop on undergraduate research*. Não citado no texto.
- Costa, A. d., Rodriguez, A. G., SIMAS, E. P. L., e Araújo, R. d. S. (2007). Lógica fuzzy: Conceitos e aplicações. Universidade do Vale do Rio dos Sinos. Não citado no texto.
- De Nicola, A., Missikoff, M., e Navigli, R. (2009). A software engineering approach to ontology building. *Information Systems*, 34:258–275. Não citado no texto.
- Ebecken, N., Lopes, M. C., Costa, M., e Rezende, S. (2005). *Sistemas Inteligentes: fundamentos e aplicações*, chapter Mineração de Textos, páginas 337–370. Manole. Não citado no texto.
- Everitt, B. S., Landau, S., e Leese, M. (2001). *Cluster Analysis*. Arnold Publishers. Não citado no texto.
- Faria, J. C. (2011). Resources of tinn-r gui/editor for r environment. UESC. Não citado no texto.
- Feldman, R. e Sanger, J. (2006). *The Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data*. Cambridge University Press. Não citado no texto.

- Fensel, D. (2000). The semantic web and its languages. *IEEE Intelligent Systems*, 15:67–73. Não citado no texto.
- Fernández-López, M. (1999). Building a chemical ontology using methodology and the ontology design environment. *IEEE Intelligent Systems*, 1:37–46. Não citado no texto.
- Frawley, W. J., Piatetsky-Shapiro, G., e Matheus, C. J. (1992). Knowledge discovery in databases: An overview. *The American Association for Artificial Intelligence (AAAI Magazine)*, 13(3):213–228. Não citado no texto.
- Freitas, F. L. G. (2006). Ontologias e a web semântica. Universidade Católica de Santos - UniSantos. Não citado no texto.
- Gómez-Pérez, A. (1999). Ontological engineering: A state of the art. *Expert Update British Computer Society*, 2:33–43. Não citado no texto.
- Gonçalves, A. P. (2007). Aplicação de lógica fuzzy em guerra eletrônica. Instituto Tecnológico de Aeronáutica. Não citado no texto.
- Gosset, W. (1947). Student's. Collected Papers, Biometrika Office. Não citado no texto.
- Griss, M. L., Favaro, J., e Alessandro, M. d. (1998). Integrating feature modeling with the rseb. In *Proceedings of the 5th International Conference on Software Reuse*. Não citado no texto.
- Gruber, T. R. (1992). Ontolingua: A mechanism to support portable ontologies. Relatório técnico, Stanford University, Knowledge Systems Laboratory. Não citado no texto.
- Gruninger, M. e Fox, M. S. (1995). Methodology for the design and evaluation of ontologies. In *Workshop on basic ontological issues in knowledge sharing*. Não citado no texto.
- Guarino, N. (1997). Understanding, building, and using ontologies. *International Journal of Human and Computer Studies*, 46:293–310. Não citado no texto.
- Guimarães, F. J. Z. (2002). Utilização de ontologias no domínio b2c. Dissertação de Mestrado, Pontifícia Universidade Católica do Rio de Janeiro - Departamento de Informática. Não citado no texto.

- Haav, H. M. e Lubi, T. L. (2001). A survey of concept-based information retrieval tools on the web. In *East- European Conference ADBIS*. Não citado no texto.
- Han, J. e Kamber, M. (2006). *Data Mining: Concepts and Techniques*. Morgan Kaufmann. Não citado no texto.
- He, S. e Fang, M. (2008). Ontological user profiling on personalized recommendation in e-commerce. *IEEE International Conference on eBusiness Engineering*, 0:585–589. Não citado no texto.
- Herlocker, J. (2000). *Understanding and Improving automated collaborative filtering systems*. Tese de Doutorado, University of Minnesota. Não citado no texto.
- Herlocker, J., Konstan, J., e Riedl, J. (2000). Explaining collaborative filtering recommendations. *ACM 2000 Conference on Computer Supported Cooperative Work*, 1:241–250. Não citado no texto.
- Hudelot, C., Atif, J., e Bloch, I. (2008). Fuzzy spatial relation ontology for image interpretation. *Fuzzy Sets and Systems*, páginas 1929–1951. Não citado no texto.
- Ivanova, T., Terzieva, V., e Andreev, R. (2010). Ontology-based recommendation system for personalized education of dyslexics to read. *Proceedings of International Conference on Software, Services & Semantic Technologies*, página 68. Não citado no texto.
- Jacobson, I., Booch, G., e Rumbaugh, J. (1999). *The unified software development process*. Addison-Wesley Longman Publishing Co., Inc. Não citado no texto.
- Jain, A. K., Murty, M., e Flynn, P. (1999). *Data clustering: a review*. ACM Computing Surveys. Não citado no texto.
- Jané, D. d. A. (2004). Uma introdução ao estudo da lógica fuzzy. *Hórus - Revista de Humanidades e Ciências Sociais Aplicadas*. Não citado no texto.
- Jasper, R. e Uschold, M. (1999). A framework for understanding and classifying ontology applications. In *IJCAI-99, Ontology Workshop*. Não citado no texto.

- Jiang, X. e Tan, A.-H. (2006). Learning and inferencing in user ontology for personalized semantic web services. In *Proceedings of the 15th international conference on World Wide Web*, páginas 1067–1068. Não citado no texto.
- Khan, L., McLeod, D., e Hovy, E. (2004). Retrieval effectiveness of an ontology-based model for information selection. *The VLDB Journal*, 13:71–85. Não citado no texto.
- Koren, Y., Bell, R., e Volinsky, C. (2009). Matrix factorization techniques for recommender systems. *IEEE Computer*, 42:30–37. Não citado no texto.
- Lang, K. (1995). Newsweeder: Learning to filter news. In *12th International Conference on Machine Learning*, páginas 331–339. Não citado no texto.
- Lee, C. S., Jian, Z. W., e Huang, L. K. (2005). A fuzzy ontology and its application to news summarization. *IEEE Transactions on Systems, Man, and Cybernetics*, 35:859–880. Não citado no texto.
- Lenat, D. e Guha, R. (1990). *Building large knowledge-based systems*. Addison-Wesley. Não citado no texto.
- Linden, F. J. v. d., Schmid, K., e Rommes, E. (2007). *Software Product Lines in Action: The Best Industrial Practice in Product Line Engineering*. Springer-Verlag New York, Inc. Não citado no texto.
- Liu, L., Kang, J., Yu, J., e Wang, Z. (2005). A comparative study on unsupervised feature selection methods for text clustering. In *International Conference on Natural Language Processing and Knowledge Engineering (NLP-KE)*, páginas 597–601. Não citado no texto.
- Lops, P., Gemmis, M., e Semeraro, G. (2011). *Content-based Recommender Systems: State of the Art and Trends*, chapter 3, páginas 73–105. Springer US. Não citado no texto.
- Maedche, A. e Staab, S. (2001). Ontology learning for the semantic web. *IEEE Intelligent Systems*, 16:72–79. Não citado no texto.
- Mahmood, T. e Ricci, F. (2009). Improving recommender systems with adaptive conversational strategies. *Proceedings of the 20th ACM conference on Hypertext and hypermedia*, páginas 73–82. Não citado no texto.
- Manning, C. D., Raghavan, P., e Schütze, H. (2008). *An Introduction to Information Retrieval*. Cambridge University Press. Não citado no texto.

- Middleton, S. E., Shadbolt, N. R., e De Roure, D. C. (2004). Ontological user profiling in recommender systems. *ACM Transactions on Information Systems*, 22:54–88. Não citado no texto.
- Missikoff, M. e Taglino, F. (2004). An ontology-based platform for semantic interoperability. In *Handbook on Ontologies*. Não citado no texto.
- Mizoguchi, R., Vanwelkrnhuysen, J., e Ikeda, M. (1994). Task ontology for reuse of problem solvin knowledge. In *ECAI'94 Towards very large knowledge bases*, páginas 46–59. Não citado no texto.
- Mohammed, N., Duong, T., e Jo, G. (2010). Contextual information search based on ontological user profile. In *Computational Collective Intelligence. Technologies and Applications*, volume 6422, chapter 52, páginas 490–500. Springer Berlin / Heidelberg. Não citado no texto.
- Montgomery, D. (1999). *Design and Analysis of Experiments*. John Wiley & Sons Australia, Limited. Não citado no texto.
- Moraes, C. M. M. e Marley, V. (2007). Fuzzycom - componente de lógica fuzzy. Relatório técnico, Pontificia Universidade Católica do Rio de Janeiro - Departamento de Engenharia Elétrica. Não citado no texto.
- Morrissey, J. (1990). Imprecise information and uncertainty in information systems. *ACM Transactions on Information Systems (TOIS)*, 8:159–180. Não citado no texto.
- Moura, M. F. (2006). *Uma abordagem para a construção e atualização de taxonomias de tópicos a partir de coleções de textos dinâmicas*. Tese de Doutorado, Instituto de Ciências Matemáticas e de Computação (ICMC - USP). Não citado no texto.
- Pagliosa, A. L. (2006). Obtenção das funções de pertinência de um sistema neurofuzzy modificado pela rede de kohonen. Dissertação de Mestrado, Universidade do Estado de Santa Catarina - UDESC. Não citado no texto.
- Parry, D. (2004). Fuzzification of a standard ontology to encourage reuse. In *Proceedings of the 2004 IEEE International Conference on Proceedings of the 2004 IEEE International Conference on*, páginas 582 – 587. Não citado no texto.

- Pinho, A. F. (1999). Uma contribuição para a resolução de problemas de programação de operações em sistemas de produção intermitentes flow-shop: A consideração de incertezas. Dissertação de Mestrado, Universidade Federal de Itajubá. Não citado no texto.
- Quan, T., Hui, S., e Cao, T. (2004). FOGA: A Fuzzy Ontology Generation Framework for Scholarly Semantic Web. In *Knowledge Discovery and Ontologies Workshop (KDO)*. Não citado no texto.
- Resnick, P. e Varian, H. R. (1997). Recommender systems. *Communications of the ACM*, 40:55–58. Não citado no texto.
- Rezende, S. O. (2003). *Sistemas Inteligentes - Fundamentos e Aplicações*. Manole, Barueri. Não citado no texto.
- Ricci, F., Rokach, L., e Shapira, B. (2011). *Introduction to Recommender Systems Handbook*. Springer. Não citado no texto.
- Ross, D. (1977). Structured analysis (sa): A language for communicating ideas. *IEEE Transactions on Software Engineering*, 3:16–34. Não citado no texto.
- Rossi, R. G. (2011). Representação de coleções de documentos textuais por meio de regras de associação. Dissertação de Mestrado, Instituto de Ciências Matemáticas e de Computação (ICMC-USP). Não citado no texto.
- Salton, G. (1989). *Automatic text processing: the transformation, analysis, and retrieval of information by computer*. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA. Não citado no texto.
- Salton, G. e Buckley, C. (1988). Term-weighting approaches in automatic text retrieval. In *INFORMATION PROCESSING AND MANAGEMENT*, páginas 513–523. Não citado no texto.
- Sandri, S. e Correa, C. (1999). Lógica nebulosa. In *V Escola de Redes Neurais, Conselho Nacional de Redes Neurais*, páginas 73–90, São José dos Campos - SP. Não citado no texto.
- Satler, M. F., Romero, F. P., Menéndez, V. H., Zapata, A., e Prieto, M. E. (2010). A fuzzy ontology approach to represent user profiles in e-learning environments. In *FUZZ-IEEE*, páginas 1–8. IEEE Computer Society. Não citado no texto.

- Schafer, J., Konstan, J., e Riedl, J. (1999). Recommender systems in e-commerce. In *ACM Conference on E-Commerce*, páginas 158–166. Não citado no texto.
- Schafer, J., Konstan, J., e Riedl, J. (2001). E-commerce recommendation applications. *Data Mining and Knowledge Discovery*, 5:115–153. Não citado no texto.
- Shardanand, U. e Maes, P. (1995). Social information filtering: Algorithms for automating 'word of mouth'. In *Conference Human Factors in Computing Systems*. Não citado no texto.
- Shaw, I. e Simões, M. G. (1998). As vantagens da aplicação da lógica difusa a sistemas de controle. In *Eletricidade Moderna*. Não citado no texto.
- Sieg, A., Mobasher, B., e Burke, R. (2007). Web search personalization with ontological user profiles. *Proceedings of the sixteenth ACM conference on Conference on information and knowledge management*, 1:525–534. Não citado no texto.
- Silva, F. W. e Cazella, S. (2005). Star: Um framework para recomendação de artigos científicos baseado na relevância da opinião dos usuários e em filtragem colaborativa. In *Encontro Nacional de Inteligência Artificial (ENIA)*. Não citado no texto.
- Sneath, P. (1957). The application of computers to taxonomy. *Journal of General Micro-biology*, 1:201–226. Não citado no texto.
- Soares, M. V., Prati, R. C., e Monard, M. C. (2008). Pretext ii: Descrição da reestruturação da ferramenta de pré-processamento de textos. Technical Report 333, ICMC-USP, São Carlos - SP. Não citado no texto.
- Sokal, R. R. e Michener, C. (1958). A statistical method for evaluating systematic relationships. *University of Kansas Scientific Bulletin*, 1:1409–1438. Não citado no texto.
- Sorensen, T. (1948). A method of establishing groups of equal amplitude in plant sociology based on similarity of species content and its application to analyses of the vegetation on danish commons. *Biologiske Skrifter*, 5:35–43. Não citado no texto.

- Souza, B. d. F. M. (2011). Modelos de fatoração matricial para recomendação de vídeos. Dissertação de Mestrado, Pontifícia Universidade Católica do Rio de Janeiro. Não citado no texto.
- Souza, D. F., Urquizo, A., Lopez, W., e Prado, A. F. (2011). Amrp: Ambiente para o monitoramento remoto de pacientes. *Centro Latinoamericano de Estudios en Informática*. Não citado no texto.
- Sowa, J. (1999). *Knowledge Representation: Logical, Philosophical and Computational Foundations*. Não citado no texto.
- Staab, S., Studer, R., Schnurr, H.-P., e Sure, Y. (2001). Knowledge processes and ontologies. *IEEE Intelligent Systems*, 16:26–34. Não citado no texto.
- Su, K., Wang, H., e Pu, F. (2010). The semantic similarity of the geoinformation concept calculation model based on ontology. In *Geoinformatics*, páginas 1–5. IEEE. Não citado no texto.
- Swartout, B., Patil, R., Knight, K., e Russ, T. (1997). Toward distributed use of large-scale ontologies. *Ontological Engineering, AAAI-97 Spring Symposium Series*, 1:138–148. Não citado no texto.
- Tan, P., Steinbach, M., e Kumar, V. (2005). *Introduction to Data Mining*. Addison-Wesley Longman Publishing. Não citado no texto.
- Todorov, K., Geibel, P., e Hudelot, C. (2011). Building a fuzzy knowledge body for integrating domain ontologies. In *International Workshop on Uncertainty Reasoning for the Semantic Web (URSW)*, páginas 3–14. Não citado no texto.
- Travassos, G., Gurov, D., e Amaral, E. (2002). Introdução à engenharia de software experimental. Relatório técnico, Programa de Engenharia de Sistemas e Computação, UFRJ. Não citado no texto.
- Urquizo, A., Santos, M. T. P., e Prado, A. F. (2012). Upfon: Unified process for building fuzzy ontology. *International Conference on Fuzzy Systems and Knowledge Discovery*. Não citado no texto.
- Uschold, M. e Gruninger, M. (1996). Ontologies: principles, methods and applications. *Knowledge Engineering Review*, 2:93–115. Não citado no texto.

- Uschold, M. e King, M. (1995). Towards a methodology for building ontologies. *Workshop on Basic Ontological Issues in Knowledge Sharing*. Não citado no texto.
- Uschold, M. e King, M. (1996). Building ontologies: towards a unified methodology. In *Annual Conference of the british computer society specialist group on expert systems*. Não citado no texto.
- Wallace, M., Alexopoulos, P., Papafragkos, I., e Vassilakis, C. (2011). Supporting user roles in ontology fuzzification. *International Workshop on Semantic Media Adaptation and Personalization*, 0:57–61. Não citado no texto.
- Widyantoro, D. H. e Yen, J. (2001). A fuzzy ontology-based abstract search engine and its user studies. In *IEEE International Conference on Fuzzy Systems*, página 129101294. Não citado no texto.
- Wohlin, C., Runeson, P., Höst, M., Ohlsson, M. C., Regnell, B., e Wesslén, A. (2000). *Experimentation in software engineering: an introduction*. Kluwer Academic Publishers. Não citado no texto.
- Yaguinuma, C. A., Santos, M. T. P., Camargo, H. A., e Nogueira, T. M. (2010). A meta-ontology approach for representing vague linguistic terms and fuzzy rules for classification in ontologies. In *Proceedings of the 2010 14th IEEE International Enterprise Distributed Object Computing Conference Workshops*, páginas 263–271. IEEE Computer Society. Não citado no texto.
- Zadeh, L. A. (1965). Fuzzy Sets. *Information and Control*, 8:338–353. Não citado no texto.
- Zhao, Y. e Karypis, G. (2002). Evaluation of hierarchical clustering algorithms for document datasets. In *Proceedings of the eleventh international conference on Information and knowledge management, CIKM '02*, páginas 515–524, New York, NY, USA. ACM. Não citado no texto.
- Zhu, J., Li, B., Wang, F., e Wang, S. (2010). Mobile ontology. *JDCTA: International Journal of Digital Content Technology and its Applications*, 4:46–54. Não citado no texto.

Formulação de Avaliação de Similaridade

Neste apêndice são mostrados os dados dos 10 dispositivos móveis associados ao primeiro dispositivo base testado. O formulário que contém 6 dispositivos base e 10 alternativas de avaliação foram entregadas para a avaliação dos usuários. Como mostrado nas Tabelas [A.1](#) e [A.2](#), o formulário entregue para os usuários, tem um Dispositivo Base no início, e mostram-se as características associadas a ele. Adicionalmente, tem-se uma lista de 10 dispositivos com suas características associadas. Na parte inferior do formulário o usuário tem que fazer uma marcação dos 7 dispositivos que ele considere o mais similares ao Dispositivo Base mostrado no início. Esses valores de marcação variam de usuário a usuário, devido a que cada um usa para marcar, as características mais representativas para ele.

As Tabelas [A.1](#) e [A.2](#) mostram a lista de características associadas ao primeiro dispositivo base apresentado para os usuários:

Dispositivo Base 01

	Disp. Base	Dispo1	Dispo2	Dispo3	Dispo4	Dispo5
	RIM OS	RIM OS	RIM OS	RIM OS	RIM OS	RIM OS
Sistema Operacional	BlackBerry	BlackBerry	BlackBerry	BlackBerry	BlackBerry	BlackBerry
Visão Geral	BlackBerry	BlackBerry	BlackBerry	BlackBerry	BlackBerry	BlackBerry
Teclado <i>Qwerty</i>	✗	✓	✓	✓	✓	✓
Conetividade	WiFi	✗	✓	✓	✗	✓
	NFC	✗	✗	✗	✗	✗
Data Transfer	GPRS	✗	✓	✓	✓	✓
	EDGE	✗	✓	✓	✓	✓
	UMTS	✗	✓	✓	✓	✓
	HSDPA	✗	✓	✓	✓	✓
	HSUPA	✗	✓	✓	✓	✓
	Video <i>Player</i>	✗	✗	✗	✗	✗
	CSS	✗	✗	✗	✗	✗
	HTML 5	html4	html4	html4	html4	html4
	Java	✗	✓	✓	✓	✓
	Protocolo Https	✓	✓	✓	✓	✓
Funções	Ringtones/Polifônico	✗	✗	✓	✗	✗
	Gravadora de áudio	✗	✗	✗	✗	✗

Tabela A.1 – Continua da página anterior

Dispositivo Base 01						
	Disp. Base	Dispo1	Dispo2	Dispo3	Dispo4	Dispo5
Câmara	×	×	×	×	×	×
	×	×	×	×	×	×
	✓	✓	✓	✓	✓	✓
	×	×	×	×	×	×
MIDP 2.0/CLDC 1.1	×	×	×	×	×	×
	×	×	×	×	×	×
Flash_Lite	✓	✓	✓	✓	✓	✓
Suporta PDF	×	×	×	×	×	×
Suporta Imagens	×	×	×	×	×	×
	✓	✓	✓	✓	✓	✓
	×	×	✓	✓	✓	✓
	✓	✓	✓	✓	✓	✓
	✓	✓	✓	✓	✓	✓
	✓	✓	✓	✓	✓	✓
Imagem	45x50	27x27	53x40	27x27	27x27	37x49
	480x360px	160x160px	320x240px	260x240px	260x240px	360x480px
	✓	×	×	×	×	×
	65536	256	65536	65536	65536	65536
	×	×	×	×	×	×
Tela	✓	✓	✓	✓	✓	✓
SMS/MMS	✓	✓	✓	✓	✓	✓
	✓	✓	✓	✓	✓	✓

Tabela A.1 – Continua da página anterior

Dispositivo Base 01

	Disp. Base	Dispo1	Dispo2	Dispo3	Dispo4	Dispo5
MMS	✘	✘	✘	✔	✘	✘
Suporta Vídeo	✔	✘	✔	✔	✔	✔
Playback Audio	✘	✘	✔	✔	✘	✔
*flv	✘	✘	✘	✘	✘	✘
*mp4	✔	✘	✔	✔	✘	✔
*wmv	✘	✘	✘	✘	✘	✘
*mov	✘	✘	✘	✘	✘	✘
*wav	✘	✘	✔	✔	✔	✔
*mmf	✘	✘	✘	✘	✘	✘
*smf	✘	✘	✘	✘	✘	✘
*mld	✘	✘	✘	✘	✘	✘
*sp_midi	✘	✘	✔	✔	✔	✔
*rmf	✘	✘	✘	✘	✘	✘
*xmf	✘	✘	✘	✘	✘	✘
*compactmidi	✘	✘	✘	✘	✘	✘
*mp3	✔	✘	✔	✔	✔	✔

Tabela A.2 – Continua da página anterior

Dispositivo Base 01

	Disp. Base	Dispo6	Dispo7	Dispo8	Dispo9	Dispo10
Câmara	✘	✘	✘	✘	✘	✘
MIDP 2.0/CLDC 1.1	✘	✘	✘	✘	✘	✘
Flash_Lite	✓	✓	✓	✓	✓	✓
Suporta PDF	✘	✘	✘	✘	✘	✘
Suporta Imagens	✘	✘	✘	✘	✘	✘
*wbmp	✓	✓	✓	✓	✓	✓
*bmp	✘	✘	✓	✓	✓	✓
*gif	✓	✓	✓	✓	✓	✓
*jpg	✓	✓	✓	✓	✓	✓
*png	✓	✓	✓	✓	✓	✓
Tamanho (mm)	45x50	108x50	27x27	40x50	27x27	37x49
Resolução	480x360px	400x360px	400x560px	240x320px	240x320px	360x480px
Tela <i>TouchScreen</i>	✓	✘	✘	✘	✘	✘
Cores	65536	256	65536	65536	65536	65536
Orientação Dupla	✘	✘	✘	✘	✘	✘
SMS/MMS	✓	✓	✓	✓	✓	✓

Tabela A.2 – Continua da página anterior

Dispositivo Base 01

	Disp. Base	Dispo6	Dispo7	Dispo8	Dispo9	Dispo10
MMS	✘	✘	✘	✘	✘	✘
Suporta Vídeo	✔	✔	✔	✔	✔	✔
Playback Audio	✘	✔	✘	✔	✔	✘
*flv	✘	✘	✘	✘	✘	✘
*mp4	✔	✔	✔	✔	✔	✔
*wmv	✘	✘	✘	✘	✘	✘
*mov	✘	✘	✘	✘	✘	✘
*wav	✘	✘	✔	✔	✔	✔
Audio/Video	✘	✘	✘	✘	✘	✘
*mmf	✘	✘	✘	✘	✘	✘
*smf	✘	✘	✘	✘	✘	✘
*mld	✘	✘	✘	✘	✘	✘
*sp_midi	✘	✘	✔	✔	✔	✔
*rmf	✘	✘	✘	✘	✘	✘
*xmf	✘	✘	✘	✘	✘	✘
*compactmidi	✘	✘	✘	✘	✘	✘
*mp3	✔	✘	✔	✘	✘	✔

Características mais representativas de cada dispositivo

Para o experimento foram consideradas as características mais representativas dos dispositivos móveis. Essas características, obtidas do WURFL¹, foram consideradas para os processos de avaliação e são detalhadas a continuação:

Visão Geral : Neste conjunto estão agrupadas as características gerais dos dispositivos móveis.

- Sistema Operacional: dá informação sobre o Sistema operativo do dispositivo.
- *Browser*: dá informação sobre o explorador do dispositivo, como *Openware* Nokia, Opera, ACESS, Teleca, entre outros.
- Teclado *Qwerty*: indica se o dispositivo móvel tem incluso um teclado. Esta característica é booleana.

Funções : Neste conjunto estão agrupadas algumas das características funcionais dos dispositivos.

- Vídeo *Player*: esta característica booleana, indica se o dispositivo suporta a reprodução de vídeos no *browser*.

¹http://wurfl.sourceforge.net/help_doc.php

- CSS: se esta característica está ativa permite colorir as tabelas quando o cor foi definido usando um arquivo CSS. Usado principalmente na tela de fundo do dispositivo.
- HTML5: indica o melhor DTD (*Document Type Definition*) para o dispositivo.
- Java: pode-se dizer que um dispositivo permite *Javascript* se suporta as características de: alerta, confirmação, *setTimeout*, *setInterval*, *document.location*.
- Protocolo HTTPS: brinda um suporte genérico para o protocolo HTTPS (conexões SSL).
- *Ringtones* / Polifônico: brinda suporte para arquivos “.midi” polifônicos.
- Gravadora de áudio: o dispositivo possui um gravador de áudio embutido.
- Câmera: o dispositivo possui uma câmera embutida.
- MIDP 2.0 / CLDC 1.1: o valor é verdadeiro se o dispositivo é compatível com MIDP (*Mobile Information Device Profile*) 2.0. Essa característica é uma especificação para o uso da plataforma Java só que aplicada para dispositivos móveis como o J2ME.
- *Flash_Lite*: característica booleana, que indica se o dispositivo suporta aplicações em flash.
- Suporte PDF: indica se o dispositivo está habilitado para ler arquivos PDF.

Data Transfer : Neste conjunto, agrupam-se as características de rede do dispositivo, as quais indicam a largura máxima de banda alcançável pelo dispositivo. As características deste conjunto são booleanas, de ter ativa uma característica, utiliza-se a transferência de banda associada a mesma.

- GPRS: 40kb de transmissão.
- EDGE: 200kb de transmissão.
- UMTS: 384kb de transmissão.
- HSDPA: entre 1800, 3600, 7200 e 14400kb de transmissão.
- HSUPA: acima de 14400kb de transmissão.

Tela : Neste conjunto, agrupam-se as características referentes com a tela do dispositivo.

- Tamanho (mm): representa a largura e altura do dispositivo em milímetros (mm).
- Resolução: representa a largura e altura da tela do dispositivo.
- *TouchScreen*: indica se o dispositivo responde aos comandos através de interações tátil.
- Cores: representa o número de cores usados pelo dispositivo.
- Orientação Dupla: indica se o dispositivo pode ser invertido, ou seja, o usuário pode mudar a orientação de largura e altura do dispositivo para navegação na web e, possivelmente, para outras funções.

Conectividade : Neste conjunto, estão agrupadas as características que permitem o estabelecimento de comunicação entre dispositivos.

- WiFi: indica se o dispositivo pode aceder a uma conexão wifi.
- NFC: indica se o dispositivo suporta NFC (*Near Field Communication*), que permite transações simplificadas de troca de dados e conexões sem fio entre dois dispositivos próximos um a outro.

Imagem : Este grupo está conformado pelas características de imagem associadas ao dispositivo.

- *Download* *imagens: indica se o dispositivo é capaz de baixar imagens.
- Formato de Imagem *wbmp: indica o suporte a imagens do tipo .wbmp.
- Formato de Imagem *bmp: indica o suporte a imagens do tipo .bmp.
- Formato de Imagem *gif: indica o suporte a imagens do tipo .gif.
- Formato de Imagem *jpg: indica o suporte a imagens do tipo .jpg.
- Formato de Imagem *png: indica o suporte a imagens do tipo .png.

SSM - MMS : Neste grupo, são consideradas as características de comunicação através de mensagens de texto simples ou multimédia.

- SMS: indica se o dispositivo pode enviar mensagens de texto plano.

- MMS: indica se o dispositivo tem a capacidade de enviar ou receber mensagens de texto multimédia.

Áudio / Vídeo : Neste conjunto, consideram-se as características de áudio e vídeo associadas ao dispositivo.

- Suporta vídeo: indica se o dispositivo pode baixar ou reproduzir arquivos de vídeo.
- *Playback* áudio: indica se o dispositivo é capaz de reproduzir arquivos de áudio o vídeo enquanto eles estão ainda sendo baixados da internet.
- Formato vídeo *flv: indica o suporte à reprodução de vídeos do tipo .flv.
- Formato vídeo *mp4: indica o suporte à reprodução de vídeos do tipo .mp4.
- Formato vídeo *wmv: indica o suporte à reprodução de vídeos do tipo .wmv.
- Formato vídeo *mov: indica o suporte à reprodução de vídeos do tipo .mov.
- Formato de sonido *wav: indica o suporte à reprodução de áudio do tipo .wav.
- Formato de sonido *mmf: indica o suporte à reprodução de áudio do tipo .mmf.
- Formato de sonido *smf: indica o suporte à reprodução de áudio do tipo .smf.
- Formato de sonido *midi: indica o suporte à reprodução de áudio do tipo .midi.
- Formato de sonido *sp_midi: indica o suporte à reprodução de áudio do tipo .sp_midi.
- Formato de sonido *rmf: indica o suporte à reprodução de áudio do tipo .rmf.
- Formato de sonido *xmf: indica o suporte à reprodução de áudio do tipo .xmf.
- Formato de sonido *compactmidi: indica o suporte à reprodução de áudio do tipo .compactmidi.

- Formato de sonido *mp3: indica o suporte à reprodução de áudio do tipo .mp3.

Valores de Erro Abosulto (MAE)

	User01		User02		User03		User04		User05	
	Onto.	Agrup.	Onto.	Agrup.	Onto.	Agrup.	Onto.	Agrup.	Onto.	Agrup.
Cjto. 01	0.8	3.0	2.2	3.4	2.2	3.8	0.4	3.0	1.6	3.0
Cjto. 02	1.6	3.8	1.6	3.8	2.0	3.8	1.8	3.8	2.2	3.8
Cjto. 03	1.6	1.8	1.6	1.0	1.6	1.4	1.8	2.6	1.0	1.8
Cjto. 04	1.2	2.2	1.2	2.2	1.6	2.2	1.2	3.2	1.4	2.0
Cjto. 05	1.2	2.8	1.2	3.2	1.6	3.2	2.2	3.8	2.8	4.0
Cjto. 06	0.8	1.0	0.4	0.6	0.6	1.0	2.2	2.0	1.6	1.4

	User06		User07		User08		User09		User10	
	Onto.	Agrup.	Onto.	Agrup.	Onto.	Agrup.	Onto.	Agrup.	Onto.	Agrup.
Cjto. 01	2.2	3.8	1.8	2.6	1.8	3.8	1.8	2.6	1.0	3.0
Cjto. 02	1.8	3.8	2.2	4.2	2.2	3.8	1.8	3.8	1.8	3.8
Cjto. 03	1.6	2.2	1.6	1.8	1.2	2.8	1.2	2.0	1.0	1.8
Cjto. 04	1.2	2.2	1.8	3.0	1.4	2.0	1.6	2.2	1.2	2.2
Cjto. 05	2.0	4.0	1.0	2.8	2.4	4.0	2.0	4.0	2.0	3.6
Cjto. 06	1.6	1.4	0.8	1.0	2.0	2.4	1.4	1.2	1.6	1.4

Cálculos de Precisão e Cobertura média

Neste Apêndice são apresentados os cálculos de precisão, cobertura e F-Score obtidos para o conjunto dos 6 dispositivos avaliados. Nas tabelas [D.1](#), [D.2](#), [D.3](#), [D.4](#), [D.5](#) e [D.6](#), podem ser observados os valores médios calculados para cada dispositivo, a partir da avaliação realizada por cada um dos 10 usuários.

Tabela D.1: Cálculo da Precisão, cobertura e *F-Score* para o Dispositivo1 segundo as abordagens baseada em ontologia e baseada em agrupamento

DISPOSITIVO 1						
	Abordagem Ontologia			Abordagem Agrupamento		
	Precisão	Cobertura	<i>F-Score</i>	Precisão	Cobertura	<i>F-Score</i>
User 1	0.71	1	0.83	0.29	0.4	0.33
User 2	0.43	0.6	0.5	0.43	0.6	0.5
User 3	0.43	0.6	0.5	0.43	0.6	0.5
User 4	0.57	0.8	0.67	0.29	0.4	0.33
User 5	0.57	0.8	0.67	0.29	0.4	0.33
User 6	0.57	0.8	0.67	0.29	0.4	0.33
User 7	0.57	0.8	0.67	0.29	0.4	0.33
User 8	0.71	1	0.83	0.43	0.6	0.5
User 9	0.57	0.8	0.67	0.29	0.4	0.33
User 10	0.43	0.6	0.5	0.29	0.4	0.33
Média	0.56	0.78	0.65	0.33	0.46	0.38

Tabela D.2: Cálculo da Precisão, Cobertura e *F-Score* para o Dispositivo2 segundo as abordagens baseada em ontologia e baseada em agrupamento

DISPOSITIVO 2						
	Abordagem Ontologia			Abordagem Agrupamento		
	Precisão	Cobertura	<i>F-Score</i>	Precisão	Cobertura	<i>F-Score</i>
User 1	0.57	0.8	0.67	0	0	0
User 2	0.57	0.8	0.67	0	0	0
User 3	0.57	0.8	0.67	0	0	0
User 4	0.57	0.8	0.67	0	0	0
User 5	0.57	0.8	0.67	0	0	0
User 6	0.43	0.6	0.5	0	0	0
User 7	0.29	0.4	0.33	0	0	0
User 8	0.57	0.8	0.67	0	0	0
User 9	0.71	1	0.83	0	0	0
User 10	0.43	0.6	0.5	0	0	0
Média	0.53	0.74	0.62	0	0	0

Tabela D.3: Cálculo da Precisão, Cobertura e *F-Score* para o Dispositivo3 segundo as abordagens baseada em ontologia e baseada em agrupamento

DISPOSITIVO 3						
	Abordagem Ontologia			Abordagem Agrupamento		
	Precisão	Cobertura	<i>F-Score</i>	Precisão	Cobertura	<i>F-Score</i>
User 1	0.71	1	0.83	0.43	0.6	0.5
User 2	0.71	1	0.83	0.43	0.6	0.5
User 3	0.71	1	0.83	0.43	0.6	0.5
User 4	0.71	1	0.83	0.43	0.6	0.5
User 5	0.71	1	0.83	0.43	0.6	0.5
User 6	0.71	1	0.83	0.43	0.6	0.5
User 7	0.71	1	0.83	0.43	0.6	0.5
User 8	0.71	1	0.83	0.43	0.6	0.5
User 9	0.71	1	0.83	0.43	0.6	0.5
User 10	0.71	1	0.83	0.43	0.6	0.5
Média	0.71	1	0.83	0.43	0.6	0.5

Tabela D.4: Cálculo da Precisão, Cobertura e *F-Score* para o Dispositivo4 segundo as abordagens baseada em ontologia e baseada em agrupamento

DISPOSITIVO 4						
	Abordagem Ontologia			Abordagem Agrupamento		
	Precisão	Cobertura	<i>F-Score</i>	Precisão	Cobertura	<i>F-Score</i>
User 1	0.71	1	0.83	0.29	0.4	0.33
User 2	0.57	0.8	0.67	0.43	0.6	0.5
User 3	0.57	0.8	0.67	0.43	0.6	0.5
User 4	0.43	0.6	0.5	0.29	0.4	0.33
User 5	0.71	1	0.83	0.29	0.4	0.33
User 6	0.57	0.8	0.67	0.43	0.6	0.5
User 7	0.57	0.8	0.67	0.14	0.2	0.17
User 8	0.43	0.6	0.5	0.29	0.4	0.33
User 9	0.57	0.8	0.67	0.29	0.4	0.33
User 10	0.57	0.8	0.67	0.43	0.6	0.5
Média	0.57	0.8	0.67	0.33	0.46	0.38

Tabela D.5: Cálculo da Precisão, Cobertura e *F-Score* para o Dispositivo5 segundo as abordagens baseada em ontologia e baseada em agrupamento

DISPOSITIVO 5						
	Abordagem Ontologia			Abordagem Agrupamento		
	Precisão	Cobertura	<i>F-Score</i>	Precisão	Cobertura	<i>F-Score</i>
User 1	0.57	0.8	0.67	0.14	0.2	0.17
User 2	0.71	1	0.83	0.14	0.2	0.17
User 3	0.71	1	0.83	0.14	0.2	0.17
User 4	0.43	0.6	0.5	0.14	0.2	0.17
User 5	0.57	0.8	0.67	0.14	0.2	0.17
User 6	0.57	0.8	0.67	0.14	0.2	0.17
User 7	0.71	1	0.83	0.14	0.2	0.17
User 8	0.43	0.6	0.5	0.14	0.2	0.17
User 9	0.57	0.8	0.67	0.14	0.2	0.17
User 10	0.43	0.6	0.5	0.14	0.2	0.17
Média	0.57	0.8	0.67	0.14	0.2	0.17

Tabela D.6: Cálculo da Precisão, Cobertura e *F-Score* para o Dispositivo6 segundo as abordagens baseada em ontologia e baseada em agrupamento

DISPOSITIVO 6						
	Abordagem Ontologia			Abordagem Agrupamento		
	Precisão	Cobertura	<i>F-Score</i>	Precisão	Cobertura	<i>F-Score</i>
User 1	0.57	0.8	0.67	0.71	1	0.83
User 2	0.57	0.8	0.67	0.57	0.8	0.67
User 3	0.57	0.8	0.67	0.57	0.8	0.67
User 4	0.57	0.8	0.67	0.57	0.8	0.67
User 5	0.57	0.8	0.67	0.57	0.8	0.67
User 6	0.71	1	0.83	0.71	1	0.83
User 7	0.57	0.8	0.67	0.71	1	0.83
User 8	0.43	0.6	0.5	0.43	0.6	0.5
User 9	0.57	0.8	0.67	0.71	1	0.83
User 10	0.43	0.6	0.5	0.57	0.8	0.67
Média	0.56	0.78	0.65	0.61	0.86	0.72

Cálculo e Representação gráfica da Precisão, cobertura e F -score para cada abordagem

Neste apêndice são apresentados os cálculos de precisão, cobertura e f -score para os 10 usuários, considerando a abordagem de recomendação baseado em ontologia e a abordagem baseada em agrupamento de dados. Como pode ser observado, nas Tabelas [E.1](#), [E.2](#), [E.3](#), [E.4](#), [E.5](#), [E.6](#), [E.7](#), [E.8](#), [E.9](#), [E.10](#), é ilustrado as medidas de avaliação aplicadas para cada abordagem considerando um *cutt-off* de 1 até 7. O *cutt-off* foi estabelecido dessa forma para mostrar o comportamento das medidas de avaliação para cada Usuário.

Nas Figuras [E.1](#), [E.2](#), [E.3](#), [E.4](#), [E.5](#), [E.6](#), [E.7](#), [E.8](#), [E.9](#) e [E.10](#) é ilustrado graficamente o comportamento das medidas de avaliação aplicadas para as duas abordagens de Recomendação. Como observado, para todos os usuários a abordagem de recomendação baseada em ontologias apresenta melhores resultados que a abordagem de recomendação baseada em agrupamento.

Tabela E.1: Cálculos de Precisão, cobertura e F-score para o Usuário 1 considerando a abordagem de recomendação com ontologia e com agrupamento

USUARIO 1									
	Abordagem com Ontologia				Abordagem com Agrupamento				
Cutt-Off	Acert.	Prec.	Cober.	F-1	Acert.	Prec.	Cober.	F-1	
1	1	1	0.2	0.33	0	0	0	0	
2	2	1	0.4	0.57	0	0	0	0	
3	3	1	0.6	0.75	1	0.33	0.2	0.25	
4	3	0.75	0.6	0.67	1	0.25	0.2	0.22	
5	3	0.6	0.6	0.6	1	0.2	0.2	0.2	
6	4	0.67	0.8	0.73	1	0.17	0.2	0.18	
7	5	0.71	1	0.83	1	0.14	0.2	0.17	

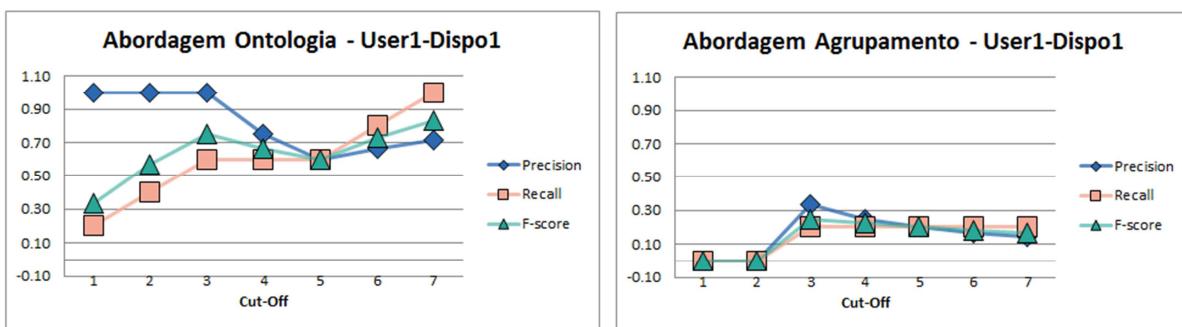


Figura E.1: Representação gráfica da precisão, cobertura e F -score do Usuário 1 considerando os *cutt-off* para a abordagem de recomendação baseada em ontologia e agrupamento

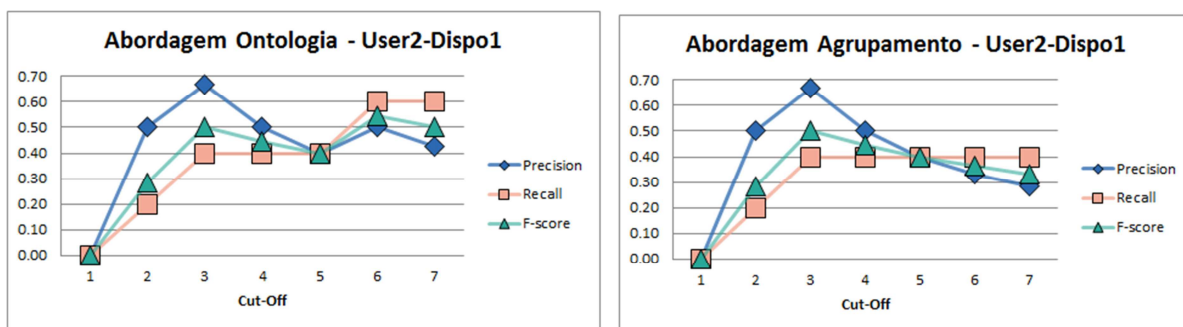


Figura E.2: Representação gráfica da precisão, cobertura e F -score do Usuário 2 considerando os *cutt-off* para a abordagem de recomendação baseada em ontologia e agrupamento

Tabela E.2: Cálculos de Precisão, cobertura e F-score para o Usuário 2 considerando a abordagem de recomendação com ontologia e com agrupamento

USUARIO 2								
	Abordagem com Ontologia				Abordagem com Agrupamento			
Cutt-Off	Acert.	Prec.	Cober.	F-1	Acert.	Prec.	Cober.	F-1
1	0	0	0	0	0	0	0	0
2	1	0.5	0.2	0.29	1	0.5	0.2	0.29
3	2	0.67	0.4	0.5	2	0.67	0.4	0.5
4	2	0.5	0.4	0.44	2	0.5	0.4	0.44
5	2	0.4	0.4	0.4	2	0.4	0.4	0.4
6	3	0.5	0.6	0.55	2	0.33	0.4	0.36
7	3	0.43	0.6	0.5	2	0.29	0.4	0.33

Tabela E.3: Cálculos de Precisão, cobertura e F-score para o Usuário 3 considerando a abordagem de recomendação com ontologia e com agrupamento

USUARIO 3								
	Abordagem com Ontologia				Abordagem com Agrupamento			
Cutt-Off	Acert.	Prec.	Cober.	F-1	Acert.	Prec.	Cober.	F-1
1	0	0	0	0	0	0	0	0
2	1	0.5	0.2	0.29	0	0	0	0
3	2	0.67	0.4	0.5	1	0.33	0.2	0.25
4	2	0.5	0.4	0.44	1	0.25	0.2	0.22
5	2	0.4	0.4	0.4	1	0.2	0.2	0.2
6	3	0.5	0.6	0.55	1	0.17	0.2	0.18
7	4	0.57	0.8	0.67	1	0.14	0.2	0.17

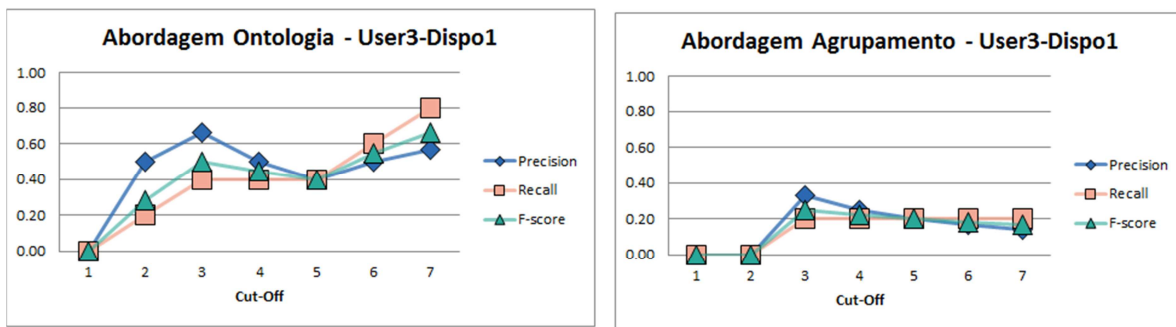


Figura E.3: Representação gráfica da precisão, cobertura e F -score do Usuário 3 considerando os *cutt-off* para a abordagem de recomendação baseada em ontologia e agrupamento

Tabela E.4: Cálculos de Precisão, cobertura e F -score para o Usuário 4 considerando a abordagem de recomendação com ontologia e com agrupamento

USUARIO 4									
	Abordagem com Ontologia				Abordagem com Agrupamento				
Cutt-Off	Acert.	Prec.	Cober.	F-1	Acert.	Prec.	Cober.	F-1	
1	1	1	0.2	0.33	1	1	0.2	0.33	
2	1	0.5	0.2	0.29	1	0.5	0.2	0.29	
3	2	0.67	0.4	0.5	1	0.33	0.2	0.25	
4	3	0.75	0.6	0.67	1	0.25	0.2	0.22	
5	4	0.8	0.8	0.8	1	0.2	0.2	0.2	
6	5	0.83	1	0.91	1	0.17	0.2	0.18	
7	5	0.71	1	0.83	1	0.14	0.2	0.17	

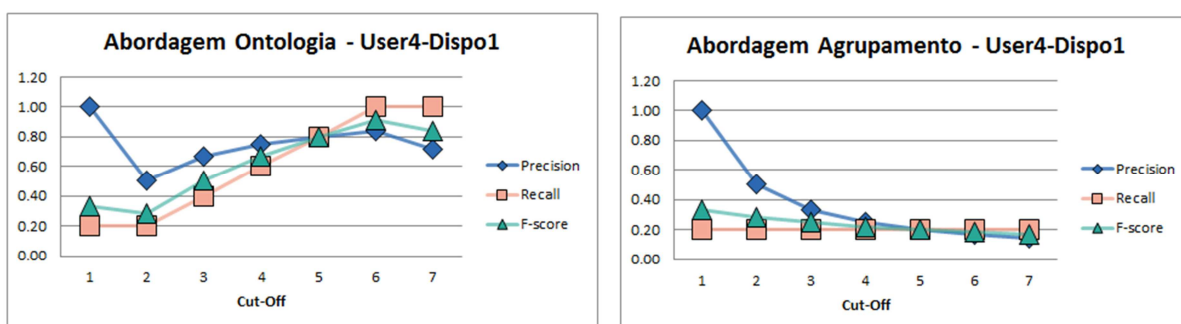


Figura E.4: Representação gráfica da precisão, cobertura e F -score do Usuário 4 considerando os *cutt-off* para a abordagem de recomendação baseada em ontologia e agrupamento

Tabela E.5: Cálculos de Precisão, cobertura e F-score para o Usuário 5 considerando a abordagem de recomendação com ontologia e com agrupamento

USUARIO 5									
	Abordagem com Ontologia				Abordagem com Agrupamento				
Cutt-Off	Acert.	Prec.	Cober.	F-1	Acert.	Prec.	Cober.	F-1	
1	0	0	0	0	1	1	0.2	0.33	
2	0	0	0	0	1	0.5	0.2	0.29	
3	1	0.33	0.2	0.25	1	0.33	0.2	0.25	
4	2	0.5	0.4	0.44	1	0.25	0.2	0.22	
5	2	0.4	0.4	0.4	1	0.2	0.2	0.2	
6	3	0.5	0.6	0.55	1	0.17	0.2	0.18	
7	4	0.57	0.8	0.67	1	0.14	0.2	0.17	

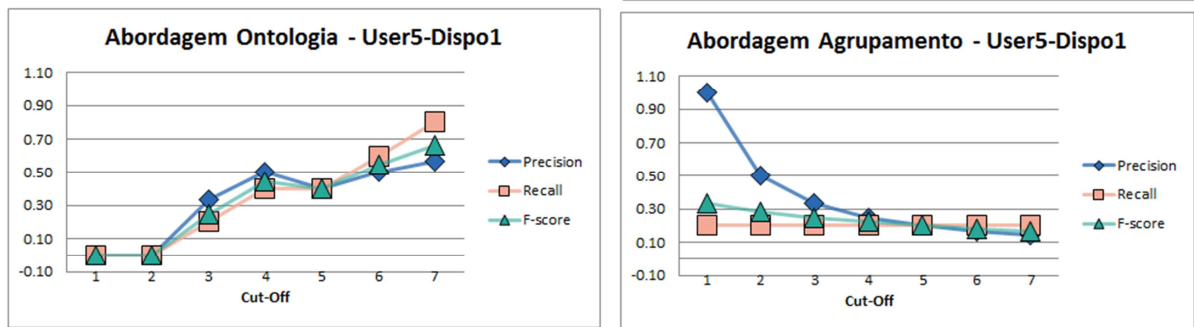


Figura E.5: Representação gráfica da precisão, cobertura e F -score do Usuário 5 considerando os $cutt-off$ para a abordagem de recomendação baseada em ontologia e agrupamento

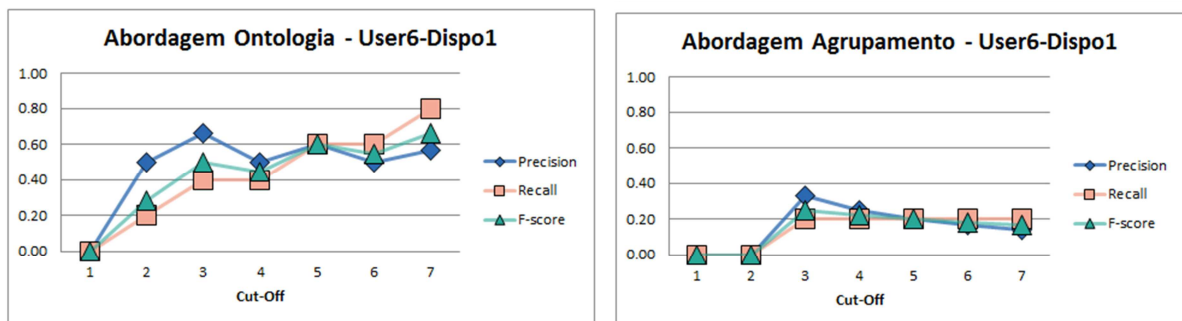


Figura E.6: Representação gráfica da precisão, cobertura e F -score do Usuário 6 considerando os $cutt-off$ para a abordagem de recomendação baseada em ontologia e agrupamento

Tabela E.6: Cálculos de Precisão, cobertura e F-score para o Usuário 6 considerando a abordagem de recomendação com ontologia e com agrupamento

USUARIO 6								
	Abordagem com Ontologia				Abordagem com Agrupamento			
Cutt-Off	Acert.	Prec.	Cober.	F-1	Acert.	Prec.	Cober.	F-1
1	0	0	0	0	0	0	0	0
2	1	0.5	0.2	0.29	0	0	0	0
3	2	0.67	0.4	0.5	1	0.33	0.2	0.25
4	2	0.5	0.4	0.44	1	0.25	0.2	0.22
5	3	0.6	0.6	0.6	1	0.2	0.2	0.2
6	3	0.5	0.6	0.55	1	0.17	0.2	0.18
7	4	0.57	0.8	0.67	1	0.14	0.2	0.17

Tabela E.7: Cálculos de Precisão, cobertura e F-score para o Usuário 7 considerando a abordagem de recomendação com ontologia e com agrupamento

USUARIO 7								
	Abordagem com Ontologia				Abordagem com Agrupamento			
Cutt-Off	Acert.	Prec.	Cober.	F-1	Acert.	Prec.	Cober.	F-1
1	1	1	0.2	0.33	1	1	0.2	0.33
2	2	1	0.4	0.57	1	0.5	0.2	0.29
3	3	1	0.6	0.75	2	0.67	0.4	0.5
4	4	1	0.8	0.89	2	0.5	0.4	0.44
5	4	0.8	0.8	0.8	2	0.4	0.4	0.4
6	4	0.67	0.8	0.73	2	0.33	0.4	0.36
7	4	0.57	0.8	0.67	2	0.29	0.4	0.33

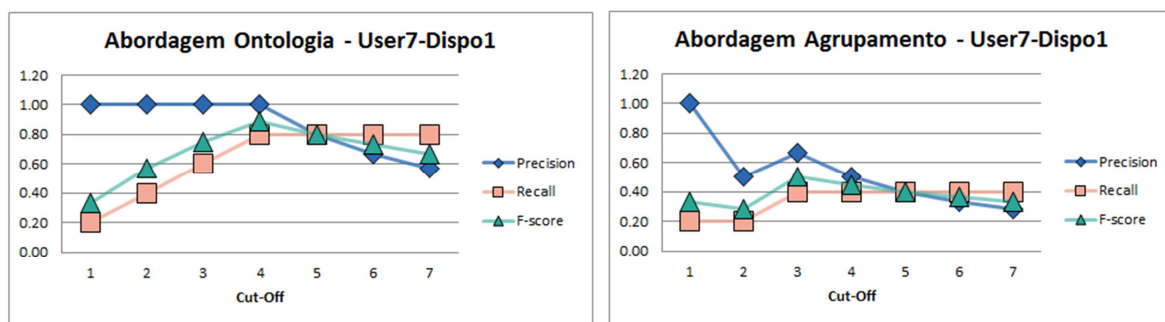


Figura E.7: Representação gráfica da precisão, cobertura e F -score do Usuário 7 considerando os *cutt-off* para a abordagem de recomendação baseada em ontologia e agrupamento

Tabela E.8: Cálculos de Precisão, cobertura e F -score para o Usuário 8 considerando a abordagem de recomendação com ontologia e com agrupamento

USUARIO 8									
	Abordagem com Ontologia				Abordagem com Agrupamento				
Cutt-Off	Acert.	Prec.	Cober.	F-1	Acert.	Prec.	Cober.	F-1	
1	1	1	0.2	0.33	0	0	0	0	
2	1	0.5	0.2	0.29	1	0.5	0.2	0.29	
3	2	0.67	0.4	0.5	1	0.33	0.2	0.25	
4	2	0.5	0.4	0.44	1	0.25	0.2	0.22	
5	3	0.6	0.6	0.6	1	0.2	0.2	0.2	
6	3	0.5	0.6	0.55	1	0.17	0.2	0.18	
7	3	0.43	0.6	0.5	1	0.14	0.2	0.17	

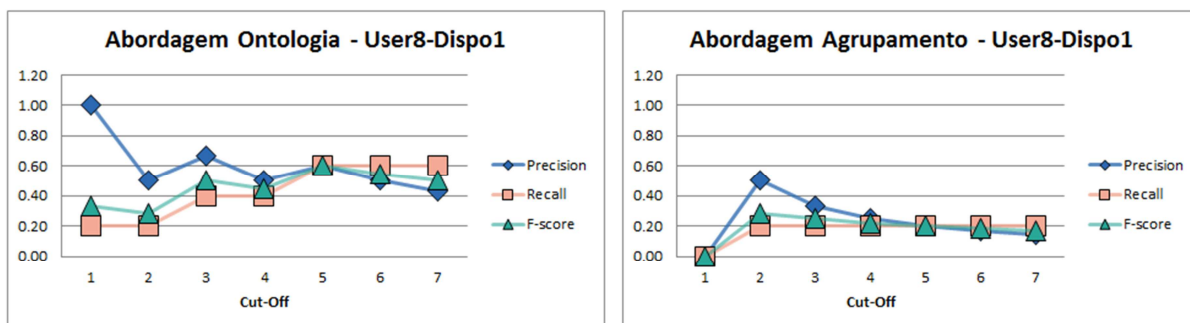


Figura E.8: Representação gráfica da precisão, cobertura e F -score do Usuário 8 considerando os *cutt-off* para a abordagem de recomendação baseada em ontologia e agrupamento

Tabela E.9: Cálculos de Precisão, cobertura e F-score para o Usuário 9 considerando a abordagem de recomendação com ontologia e com agrupamento

USUARIO 9									
	Abordagem com Ontologia				Abordagem com Agrupamento				
Cutt-Off	Acert.	Prec.	Cober.	F-1	Acert.	Prec.	Cober.	F-1	
1	0	0	0	0	0	0	0	0	
2	1	0.5	0.2	0.29	1	0.5	0.2	0.29	
3	1	0.33	0.2	0.25	2	0.67	0.4	0.5	
4	1	0.25	0.2	0.22	2	0.5	0.4	0.44	
5	2	0.4	0.4	0.4	2	0.4	0.4	0.4	
6	2	0.33	0.4	0.36	2	0.33	0.4	0.36	
7	3	0.43	0.6	0.5	2	0.29	0.4	0.33	

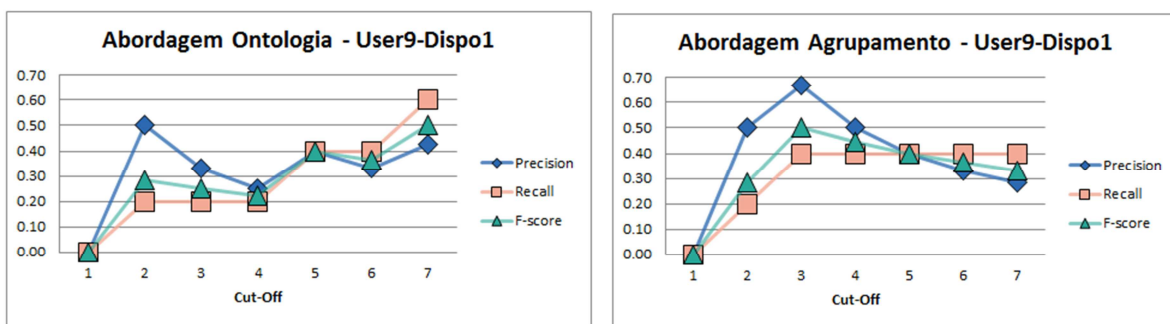


Figura E.9: Representação gráfica da precisão, cobertura e F-score do Usuário 9 considerando os *cutt-off* para a abordagem de recomendação baseada em ontologia e agrupamento

Tabela E.10: Cálculos de Precisão, cobertura e F-score para o Usuário 10 considerando a abordagem de recomendação com ontologia e com agrupamento

USUARIO 10								
	Abordagem com Ontologia				Abordagem com Agrupamento			
Cutt-Off	Acert.	Prec.	Cober.	F-1	Acert.	Prec.	Cober.	F-1
1	0	0	0	0	0	0	0	0
2	1	0.5	0.2	0.29	0	0	0	0
3	2	0.67	0.4	0.5	1	0.33	0.2	0.25
4	2	0.5	0.4	0.44	1	0.25	0.2	0.22
5	3	0.6	0.6	0.6	1	0.2	0.2	0.2
6	3	0.5	0.6	0.55	1	0.17	0.2	0.18
7	4	0.57	0.8	0.67	1	0.14	0.2	0.17

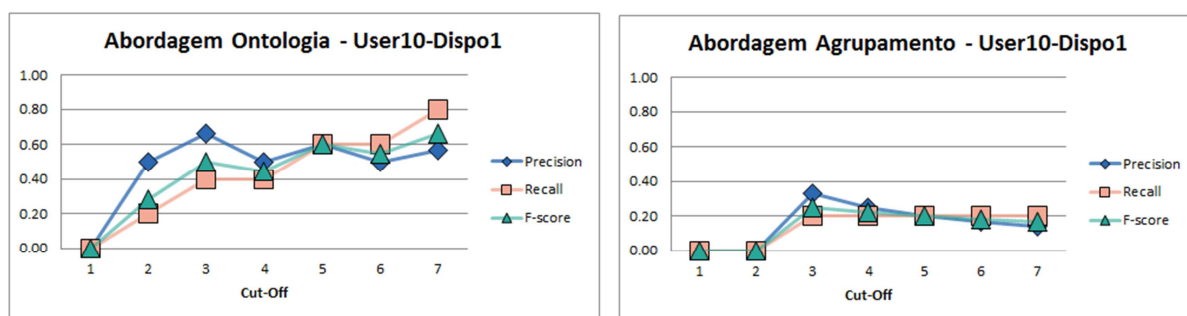


Figura E.10: Representação gráfica da precisão, cobertura e F -score do Usuário 10 considerando os *cutt-off* para a abordagem de recomendação baseada em ontologia e agrupamento

Tabela Padrão de Distribuição de Probabilidade Estatística T de Student
(Gosset, 1947)

Área contida nas duas caudas laterais da distribuição t de Student

gl/α	Graus de Significância						
Uma cauda	0.10	0.05	0.025	0.01	0.005	0.001	0.0005
Duas caudas	0.20	0.10	0.05	0.02	0.01	0.002	0.01
1	3.078	6.314	12.71	31.82	63.66	318.3	637
2	1.886	2.92	4.303	6.965	9.925	22.33	31.6
3	1.638	2.353	3.182	4.541	5.841	10.21	12.92
4	1.533	2.132	2.776	3.747	4.604	7.173	8.61
5	1.476	2.015	2.571	3.365	4.032	5.893	6.869

Continuação na página seguinte

Tabela F.1 – *Continua da página anterior***Área contida nas duas caudas laterais da distribuição t de Student**

gl/α	Graus de Significância						
	0.10	0.05	0.025	0.01	0.005	0.001	0.0005
Uma cauda	0.10	0.05	0.025	0.01	0.005	0.001	0.0005
Duas caudas	0.20	0.10	0.05	0.02	0.01	0.002	0.01
6	1.44	1.943	2.447	3.143	3.707	5.208	5.959
7	1.415	1.895	2.365	2.998	3.499	4.785	5.408
8	1.397	1.86	2.306	2.896	3.355	4.501	5.041
9	1.383	1.833	2.262	2.821	3.25	4.297	4.781
10	1.372	1.812	2.228	2.764	3.169	4.144	4.587
11	1.363	1.796	2.201	2.718	3.106	4.025	4.437
12	1.356	1.782	2.179	2.681	3.055	3.93	4.318
13	1.35	1.771	2.16	2.65	3.012	3.852	4.221
14	1.345	1.761	2.145	2.634	2.977	3.787	4.14
15	1.341	1.753	2.131	2.602	2.947	3.733	4.073
16	1.337	1.746	2.12	2.583	2.921	3.686	4.015
17	1.333	1.74	2.11	2.567	2.898	3.646	3.965
18	1.33	1.734	2.101	2.552	2.878	3.61	3.922
19	1.328	1.729	2.093	2.539	2.861	3.579	3.883
20	1.325	1.725	2.086	2.528	2.845	3.552	3.85

Continuação na página seguinte

Tabela F.1 – *Continua da página anterior*

Área contida nas duas caudas laterais da distribuição t de Student							
gl/α	Graus de Significância						
	0.10	0.05	0.025	0.01	0.005	0.001	0.0005
Uma cauda	0.20	0.10	0.05	0.02	0.01	0.002	0.01
Duas caudas							
21	1.323	1.721	2.08	2.518	2.831	3.527	3.819
22	1.321	1.717	2.074	2.508	2.819	3.505	3.792
23	1.319	1.714	2.069	2.5	2.807	3.485	3.768
24	1.318	1.711	2.064	2.492	2.797	3.467	3.745
25	1.316	1.708	2.06	2.485	2.787	3.45	3.725
26	1.315	1.706	2.056	2.479	2.779	3.435	3.707
27	1.314	1.703	2.052	2.473	2.771	3.421	3.69
28	1.313	1.701	2.048	2.467	2.763	3.408	3.674
29	1.311	1.699	2.045	2.462	2.756	3.396	3.659
30	1.31	1.697	2.042	2.457	2.75	3.385	3.646
32	1.309	1.694	2.037	2.449	2.738	3.365	3.622
34	1.307	1.691	2.032	2.441	2.728	3.348	3.601
36	1.306	1.688	2.028	2.434	2.719	3.333	3.582
38	1.304	1.686	2.024	2.439	2.712	3.319	3.566
40	1.303	1.684	2.021	2.423	2.704	3.307	3.551

Continuação na página seguinte

Tabela F.1 – Continua da página anterior

Área contida nas duas caudas laterais da distribuição t de Student							
gl/α	Graus de Significância						
	0.10	0.05	0.025	0.01	0.005	0.001	0.0005
Uma cauda	0.10	0.05	0.025	0.01	0.005	0.001	0.0005
Duas caudas	0.20	0.10	0.05	0.02	0.01	0.002	0.01
42	1.302	1.682	2.018	2.418	2.698	3.296	3.538
44	1.301	1.68	2.015	2.414	2.692	3.286	3.526
46	1.3	1.679	2.013	2.41	2.687	3.277	3.515
48	1.299	1.677	2.011	2.407	2.682	3.269	3.505
50	1.299	1.676	2.009	2.403	2.678	3.261	3.496
55	1.297	1.673	2.004	2.396	2.668	3.245	3.476
60	1.296	1.671	2	2.39	2.66	3.232	3.46
65	1.295	1.669	1.997	2.385	2.654	3.22	3.447
70	1.294	1.667	1.994	2.381	2.648	3.211	3.435
80	1.292	1.664	1.99	2.374	2.639	3.195	3.416
100	1.29	1.66	1.984	2.364	2.626	3.174	3.39
150	1.287	1.655	1.976	2.351	2.609	3.145	3.357
200	1.286	1.653	1.972	2.345	2.601	3.131	3.34