

Miriam Sayão

**Verificação e Validação em Requisitos: Processamento da
Linguagem Natural e Agentes**

Tese de Doutorado

Tese apresentada como requisito parcial para
obtenção do título de Doutor pelo Programa de Pós-
Graduação em Informática da PUC-Rio.

Orientador: Julio Cesar Sampaio do Prado Leite

Rio de Janeiro
Abril de 2007

Miriam Sayão

Verificação e Validação em Requisitos: Processamento da Linguagem Natural e Agentes

Tese apresentada como requisito parcial para obtenção do título de Doutor pelo Programa de Pós-Graduação em Informática da PUC-Rio. Aprovada pela Comissão Examinadora abaixo assinada.

prof. Julio Cesar Sampaio do Prado Leite
Orientador
PUC-Rio

prof. Marco Antonio Casanova
PUC-Rio

profa. Karin Koogan Breitman
PUC-Rio

prof. Jorge Luis Nicolas Audy
PUC-RS

profa. Vera Maria Benjamim Werneck
UERJ

prof. José Eugênio Leal
Coordenador Setorial do Centro Técnico Científico - PUC-Rio

Rio de Janeiro, 18 de abril de 2007

Todos os direitos reservados. É proibida a reprodução total ou parcial do trabalho sem autorização da universidade, da autora e do orientador.

Miriam Sayão

Graduada em Estatística pela Universidade Estadual de Campinas (UNICAMP) em 1975. Obteve o título de Mestre em Ciência da Computação pela Universidade Federal do Rio Grande do Sul (UFRGS) em 1986. Atua como docente na Pontifícia Universidade Católica do Rio Grande do Sul, na Faculdade de Informática.

Ficha Catalográfica

Sayão, Miriam

Verificação e validação em requisitos: processamento da linguagem natural e agentes / Miriam Sayão ; orientador: Julio César Sampaio do Prado Leite. – 2007.

205 f. : il. ; 30 cm

Tese (Doutorado em Informática)–Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, 2007.

Inclui bibliografia

1. Informática – Teses. 2. Análise de Requisitos. 3. Geração de visões de requisitos. 4. Agrupamento de requisitos. 5. Verificação e Validação em requisitos. 6. Omissões em requisitos não funcionais. 7. Processamento da linguagem natural em requisitos. 8. Agentes e gerenciamento de requisitos. I. Leite, Julio César Sampaio do Prado. II. Pontifícia Universidade Católica do Rio de Janeiro. Departamento de Informática. III. Título.

CDD : 004

Dedico aos meus pais, Antonio Rachid Sayão e Iracema Fischer Sacconi Sayão,
pelos ensinamentos da vida

Agradecimentos

Agradeço a todos os amigos que colaboraram de alguma forma, nesta tese. São muitos, alguns até colaboraram sem nem mesmo saber que o faziam....

A Julio César S. P. Leite, meu orientador, cujos questionamentos e revisões colaboraram para a qualidade deste trabalho.

Aos funcionários da PUC-Rio e do DI, em especial à Ruth, Débora, Emanuelle, Wagner e Rosane; a Zé Carlos e equipe de suporte. Aos professores Arndt von Staa, Carlos J. Pereira de Lucena, Ruy Milidiú, pelos ensinamentos. Ao Akeo Tanabe, pela imensa disposição em apoiar.

À minha família de origem: Iracema & Antonio, Ivette, Rosely & Fábio & Camila, Silvia, Yara & Celso & Marco, Filó, pelo carinho, sempre. Ao Aluízio, pelo afeto e apoio nesta jornada.

À minha "família" carioca (???): Karin, Carol, Lyrene, Roberta, Uirá, Eduardo, Karla, Pádua, Guga, Dani, Daflon, Léo, Mauren, Adéle, Silvana, Roberto,.... pela amizade e companheirismo.

Aos amigos de todos os lugares: Malu & Fleck, Sônia & Mário, Carlão, Eldo, Dai, Denise, Schüler, Olandina, Kleyna & Eduardo, Deise pela amizade que supera tempo e distâncias.

Ao Aluízio e ao Rafael, pelo apoio com a plataforma MIDAS. À Karin Breitman, pelo imenso apoio e amizade, sempre.

À Vera Lima e Jorge Audy, pelo grande incentivo nesta caminhada. À Faculdade de Informática da PUCRS, por todo o apoio recebido.

À Pontifícia Universidade Católica do Rio Grande do Sul (PUC-RS), à CAPES e à Pontifícia Universidade Católica do Rio de Janeiro (PUC-Rio) pelo financiamento e apoio financeiro concedidos, sem os quais este trabalho não poderia ter sido realizado.

Resumo

Sayão, Miriam; Leite, Julio César S. P. (orientador). **Verificação e Validação em Requisitos: Processamento da Linguagem Natural e Agentes**. Rio de Janeiro, 2007. 205p. Tese de Doutorado - Departamento de Informática, Pontifícia Universidade Católica do Rio de Janeiro.

No processo de desenvolvimento do software, atividades relacionadas ao Processo de Requisitos envolvem elicitação, modelagem, verificação e validação dos requisitos. O uso da linguagem natural no registro dos requisitos facilita a comunicação entre os participantes do processo, além de possibilitar que clientes e usuários validem requisitos sem necessitar de conhecimento extra. Por outro lado, na economia globalizada atual, o desenvolvimento de software por equipes geograficamente distribuídas está se tornando uma norma. Nesse cenário, atividades de verificação e validação de requisitos para um software de média ou alta complexidade podem envolver o tratamento de centenas ou milhares de requisitos. Com essa ordem de complexidade é importante que o engenheiro de software tenha apoio computacional para o desempenho adequado das atividades de aferição de qualidade. Neste trabalho estamos propondo uma estratégia que combina técnicas de processamento da linguagem natural (PLN) e agentes de software para apoiar as atividades de análise dos requisitos. Geramos visões textuais ou gráficas de grupos de requisitos relacionados; visões apóiam a análise de completude, a identificação de duplicidades e de dependências entre requisitos. Utilizamos técnicas de análise de conteúdo para apoiar a identificação de omissões em requisitos não funcionais. Também propomos uma estratégia para a construção ou atualização do léxico da aplicação, utilizando técnicas de PLN. Utilizamos agentes de software para implementar serviços que incorporam as estratégias referidas, e também para atuar como representantes dos participantes do projeto em desenvolvimento.

Palavras-chave

Análise de Requisitos, geração de visões de requisitos, agrupamento de requisitos, verificação e validação em requisitos, omissões em requisitos não funcionais, processamento da linguagem natural em requisitos, agentes e gerenciamento de requisitos.

Abstract

Sayão, Miriam; Leite, Julio Cesar S. P. (Advisor). **Requirements Verification and Validation: Natural Language Processing and Software Agents**. Rio de Janeiro, 2007. 205p. Ph. D. Thesis - Departamento de Informática, Pontifícia Universidade Católica do Rio de Janeiro.

In software development process, initial activities can involve requirements elicitation, modeling and analysis (verification and validation). The use of natural language in the register of the requirements facilitates the communication among stakeholders, besides offering possibilities to customers and users to validate requirements without extra knowledge. On the other hand, in the current global economy, software development for teams geographically distributed is becoming a rule. In this scenario, requirements verification and validation for medium or high complexity software can involve the treatment of hundreds or even thousand requirements. With this complexity order it is important to provide computational support for the software engineer execute quality activities. In this work we propose a strategy which combines natural language processing (NLP) techniques and software agents to support analysis activities. We have generated textual or graphical visions from groups of related requirements; visions help completeness analysis, identification of duplicities and dependences among requirements. We use content analysis techniques to support the identification of omissions in nonfunctional requirements. Also, we propose a strategy to construct the lexicon, using NLP techniques. We use software agents to implement web services that incorporate the related strategies, and also agents to act as personal assistants for stakeholders of the software project.

Keywords

Requirements analysis, requirements clustering, requirements visions, requirements verification and validation, lack of nonfunctional requirements, requirements and natural language processing, software agents and requirements management.

Sumário

1 Introdução	14
1.1. Desenvolvimento Distribuído de Software (DDS): desafios para o Processo de Requisitos	16
1.2. Agentes nas Atividades de Verificação e Validação de Requisitos no Desenvolvimento Distribuído de Software	18
1.3. Contribuições deste trabalho	24
1.4. Visão geral deste documento	26
2 Processo de Requisitos, PLN e Agentes	27
2.1. Processo de Requisitos: estruturação das atividades	27
2.2. Processo de requisitos em ambientes distribuídos de desenvolvimento	32
2.3. Análise de Requisitos: importância das atividades de verificação e validação	35
2.4. Verificação e Validação: uso de PLN e Agentes de Software	36
2.5. Alguns métodos e técnicas de PLN para apoio ao processo proposto	38
2.5.1. Part-Of-Speech Tagger (POS tagger)	39
2.5.2. Representação de documentos: abordagem bag-of-words	41
2.5.3. Similaridade entre documentos	45
2.5.4. Stemização	48
2.5.5. Concordanceador	51
2.5.6. Análise de conteúdo	52
2.5.7. Agrupamento ou clusterização de documentos	55
2.5.8. Recuperação de informações: medidas <i>recall</i> e <i>precision</i>	57
2.6. Processo de Requisitos, PLN e Agentes	58
3 Proposta para Análise de Requisitos utilizando PLN	59
3.1. Geração de visões dos requisitos	60
3.1.1. Categorização de requisitos e geração de visões	62
3.1.2. Identificação de termos para enriquecimento da taxonomia	66
3.1.3. Categorizar requisitos	69
3.1.4. Gerar visões	70
3.2. Construção do léxico da aplicação	73
3.2.1. Termos ou expressões não incluídos em dicionários	75
3.2.2. Identificação de sujeitos e objetos	77
3.3. Detecção de discrepâncias, erros e omissões em requisitos	82
3.3.1. Duplicidade em requisitos	83

3.3.2. Omissões em requisitos	85
4 Proposta de uma Arquitetura de Agentes	93
4.1. Adequação de Web services e agentes	94
4.2. Plataforma utilizada: MIDAS	95
4.3. Aplicação: características, agentes e modelos de papéis	98
4.3.1. Perfil de usuários	98
4.3.2. Uso do blackboard	99
4.3.3. Repositório de artefatos e de informações sobre projetos	99
4.3.4. Agentes: modelo de papéis	100
4.4. Uma visão da implementação	105
4.4.1. Estrutura da aplicação multi-agente	107
4.5. Interface com o usuário: processos abordados	108
4.6. Algumas considerações sobre a aplicação	110
5 Estudos de caso	114
5.1. Estudo de caso: documento Exit	114
5.1.1. Geração de visões	115
5.1.2. Construção do léxico da aplicação	120
5.1.3. Detecção de discrepâncias, erros e omissões em requisitos	122
5.2. Estudo de caso: documento Escalas	127
5.2.1. Construção do léxico da aplicação: termos ou expressões não dicionarizadas	127
5.2.2. Construção do léxico da aplicação: extração de sujeitos e objetos	128
5.2.3. Omissões em requisitos: análise de conteúdo para RNF's	129
5.3. Estudo de caso: documentos E-commerce/entrega e E-commerce/vendas	131
5.4. Estudo de caso: documento Selic	132
5.5. Avaliação dos resultados preliminares	133
6 Conclusões	135
6.1. Abordagem proposta e contribuições	135
6.1.1. Estratégias com uso de técnicas de processamento da linguagem natural	136
6.1.2. Estratégias com uso de agentes de software	139
6.2. Comparação com trabalhos relacionados	140
6.3. Limitações e trabalhos futuros	145
7 Referências Bibliográficas	147
Apêndice A Padrões para extração de sujeitos/atores: doc. Escalas	156
Apêndice B Padrões para extração de sujeitos/atores: doc. SELIC	158

Apêndice C Dicionário de Requisitos Não-Funcionais	160
Apêndice D Código dos agentes da aplicação	163
Anexo A Stopwords utilizadas	204

Lista de figuras

Figura 1 - Modelo de dependência estratégica para verificação e validação de requisitos	22
Figura 2 - Modelo SADT da Engenharia de Requisitos [Leite94]	27
Figura 3 - Processo de Requisitos [Sommerville04]	30
Figura 4 - Ciclo de elicitação e análise [Sommerville04]	31
Figura 5 - Impactos do desenvolvimento distribuído no Processo de Requisitos [Damian03a]	33
Figura 6 - Visão geral de um etiquetador morfossintático	39
Figura 7 - Curva segundo a lei de Zipf	44
Figura 8 - Curva de Zipf com os cortes propostos por Luhn	44
Figura 9 - Espaço vetorial para representação de documentos [Salton83]	46
Figura 10 - Contextos para a palavra concordanceador	51
Figura 11 - Diagrama ilustrativo das medidas <i>precision</i> e <i>recall</i> [Manning99]	58
Figura 12 - Visão geral do processo para categorização dos requisitos e geração de visões	65
Figura 13- Processo para enriquecer a taxonomia	67
Figura 14 - Sub-sentença apresentando contexto para o termo "reserva"	67
Figura 15 - Processo para categorização de requisitos	69
Figura 16 - Processo para geração das visões dos requisitos	71
Figura 17 - Modelo XTM para termos da taxonomia	72
Figura 18 - Modelo XTM para requisitos funcionais	72
Figura 19 - Modelo XTM para associações entre temas e requisitos	72
Figura 20 - Mapa visual mostrando requisitos relacionados ao tema pagamento	73
Figura 21 - Processo para extração de termos não-dicionarizados	76
Figura 22 - Processo geral para extração de símbolos	78
Figura 23 - Processo para identificação de requisitos similares	83
Figura 24 - Processo para análise de conteúdo para RNF's	87
Figura 25 - Estrutura do dicionário de RNF's para identificação de omissões	88
Figura 26 - Frequência das palavras no documento de requisitos	90
Figura 27 - Frequência das categorias e expressões no documento de requisitos	91
Figura 28 - Comparação de documentos tendo por base o dicionário de RNF's	92
Figura 29 - Arquitetura da plataforma Midas	97
Figura 30 - Responsabilidades e colaborações do agente Manager	100
Figura 31 - Responsabilidades e colaborações do agente Analisador Léxico	101
Figura 32 - Responsabilidades e colaborações do agente Construtor do Léxico	101

Figura 33 - Responsabilidades e colaborações do agente Gerador de Visões	102
Figura 34 - Responsabilidades e colaborações do agente Estatístico	102
Figura 35 - Responsabilidades e colaborações do agente Verificador	103
Figura 36 - Responsabilidades e colaborações do agente Validador	104
Figura 37 - Responsabilidades e colaborações do agente Comunicador	104
Figura 38 - Responsabilidades e colaborações do agente Observador	105
Figura 39 - Responsabilidades e colaborações do agente Rastreador	105
Figura 40 - Visão de alto nível da plataforma	106
Figura 41 - Visão do sistema multi-agente	107
Figura 43 - Interface para identificação de requisitos associados a um tema	109
Figura 44 - Diagrama de atividades para o processo de Verificação	110
Figura 45 - Visão da taxonomia e do tema reserva e requisitos associados	119
Figura 46 - O requisito funcional 17 e temas associados	120
Figura 47 - Estrutura do dicionário de RNF's para identificação de omissões	125
Figura 48 - Resultados para o documento <i>exit</i> - Sistema Exit	126
Figura 49 - Resultados para o documento escalas - Sistema Gestor de Escalas	130
Figura 50 - Comparando documentos através de RNF's	132

Lista de tabelas

Tabela 1 - Etiquetas utilizadas pelo QTAG	41
Tabela 2 - Frase etiquetada pelo QTAG	41
Tabela 3 - Estrutura genérica para uma matriz termo-documento	42
Tabela 4 - Resultados obtidos na aplicação de dois diferentes stemmers	49
Tabela 5 - Medidas utilizadas para cálculo de recall e precision	57
Tabela 6 - Símbolos do LAL, noção e impactos [Leite90]	74
Tabela 7 – Exemplo de símbolo de um léxico do tipo LAL	75
Tabela 8 - Padrões para extração de sujeitos e atores	80
Tabela 9 - Padrões para extração de objetos e recursos	80
Tabela 10 - Estado atual da implementação da ferramenta de suporte	112
Tabela 11 - Lista parcial de termos selecionados e índices	115
Tabela 12 - Termos candidatos (parcial)	117
Tabela 13 - Padrões para o termo reserva	117
Tabela 14 - Taxonomia e requisitos relacionados	118
Tabela 15 - Número de requisitos por ligações	118
Tabela 16 - Visão textual de requisitos associados ao tema reserva (parcial)	119
Tabela 17 - Termos não-dicionarizados do documento Exit (parcial)	121
Tabela 18 - Expressões registrando diferenças culturais e lingüísticas	121
Tabela 19 - Um termo e seus contextos	122
Tabela 20 - Matriz de similaridade de requisitos: exibição parcial	122
Tabela 21 - Requisitos candidatos à avaliação de duplicidade	123
Tabela 22 - Valores de recall e precision para duplicidade	123
Tabela 23 - Expressões não-dicionarizadas	127
Tabela 24 - Símbolos do tipo sujeito e objeto para o sistema de gestão de escalas	129
Tabela 25 - Contextos para símbolos do sistema de gestão de escalas	129
Tabela 26- Símbolos do tipo sujeito e objeto para o sistema Selic	133
Tabela 27 - Valores de recall e precision dos símbolos extraídos	134