

HUANDER UMBERTO MARTINS TIRONI

CRIAÇÃO SEMIAUTOMÁTICA DO ESQUEMA DE APRENDIZAGEM EM UM  
AMBIENTE COLABORATIVO SEMÂNTICO VOLTADO À APRENDIZAGEM  
ORGANIZACIONAL

Dissertação apresentada ao programa de Pós-Graduação em Informática da Pontifícia Universidade Católica do Paraná para obtenção do título de Mestre em Ciência da Computação.

Curitiba  
2015

HUANDER UMBERTO MARTINS TIRONI

CRIAÇÃO SEMIAUTOMÁTICA DO ESQUEMA DE APRENDIZAGEM EM UM  
AMBIENTE COLABORATIVO SEMÂNTICO VOLTADO À APRENDIZAGEM  
ORGANIZACIONAL

Dissertação apresentada ao Programa de Pós-Graduação em Informática da Pontifícia Universidade Católica do Paraná para obtenção do título de Mestre em Ciência da Computação.

Área de Concentração: Engenharia de Software

Orientadora: Prof. Dra. Andreia Malucelli  
Co-Orientador: Dr. André Luís Andrade Menolli

Este exemplar foi revisado e alterado em relação à versão original, sob responsabilidade única do autor, com anuência de seu orientador.

Curitiba, 10 de Abril de 2014.

Assinatura do Autor

Assinatura do Orientador

#### FICHA CATALOGRÁFICA

Tironi, Huander

Criação Semiautomática de Unidades de Aprendizagem em um Ambiente Colaborativo Semântico Voltado à Aprendizagem Organizacional

H. U. M Tironi. – Curitiba 2015.

Número de páginas p.

Dissertação (Mestrado) – Pontifícia Universidade Católica do Paraná. Curitiba. Programa de Pós-Graduação em informática.

1. Aprendizagem Organizacional 2. Objetos de Aprendizagem Organizacional 3. Unidades de Aprendizagem Organizacional 4. Engenharia

## DEDICATÓRIA

Aos meus pais Jusmar e Vilma Tironi.  
À Luciana, minha futura e amada esposa.  
Aos meus irmãos Ismaile e Heiridan.  
Ao meu grande amigo Ulisses Piassa.

## AGRADECIMENTOS

Agradeço imensamente aos meus pais, que são os meus mestres, Jusmar e Vilma, pois, me ensinaram o valor da educação, da luta e da honestidade, e foram meus portos seguros nos momentos em que precisei descansar.

À Luciana Mendes, minha noiva e companheira que amo muito, que compreendeu todo o esforço e esteve ao meu lado, quando precisei me dedicar ao trabalho. Que também me mostrou o caminho a seguir quando estive fora de foco.

Aos meus irmãos, Ismaile e Heiridan, que me acompanham na vida e me dão aula do que é trabalho duro me incentivando a lutar pelos meus sonhos.

À minha orientadora, professora e amiga Andreia Malucelli, que nesses três anos de estudo foi responsável pela minha evolução perante este desafio que é o Mestrado, no qual refletiu na forma como eu encaro os desafios que são postos a mim.

Ao meu coorientador, professor e amigo André Menolli, que gentilmente cedeu seu tempo para explicar, orientar e acompanhar o desenvolvimento desta pesquisa, e juntamente com sua namorada, Kassia, abriram as portas de sua casa para que a pesquisa pudesse ser discutida e evoluída.

À professora Sheila Reinehr, por sua empatia, apoio e atenção, e amizade, com os quais colaborou com o meu crescimento educacional e profissional, e que também esteve disponível quando precisei de sua orientação.

Ao meu amigo, Ulisses Piassa, com o qual pude compartilhar as dificuldades e conquistas durante estes anos de Mestrado. A todos os colegas e amigos de profissão, contribuindo para a realização deste trabalho com a participação dos experimentos.

Finalmente, agradeço a Deus, que me deu o sonho de conquistar este Mestrado, o temor que me moveu durante estes anos, e a vontade para alcançar este objetivo. Que também me protegeu em todas as viagens e delineou todas as etapas desde o início, que me fizeram chegar à etapa final deste mestrado.

“Você pode encontrar as coisas que perdeu, mas  
nunca as coisas que abandonou.”  
(J.R.R. Tolkien)

## RESUMO

A aprendizagem organizacional é uma área que auxilia as organizações a melhorar seus processos por meio de reuso de experiências, fazendo com que o conhecimento seja acessível a toda a organização. No entanto, não é uma tarefa trivial conseguir que haja aprendizagem em empresas de desenvolvimento de software, principalmente, por ser uma área em que os processos e conhecimentos estão muitas vezes internalizados na mente de seus funcionários. A literatura apresenta vários trabalhos que buscam meios de facilitar e incentivar o compartilhamento de informações e a disseminação do conhecimento. Dentre estes trabalhos está um Ambiente Colaborativo Semântico, que tem como objetivo organizar a informação gerada pelos desenvolvedores, por meio de uma ferramenta social (wiki), e assim, com a ajuda de um especialista que seja capaz de implementar esquemas de aprendizagem, criar unidades de aprendizagem que serão tratadas como cursos, lições ou qualquer parte limitada de um assunto específico. Entretanto, a dependência de um especialista na fase de criação de esquemas de aprendizagem trouxe uma série de limitações para esta abordagem, como por exemplo, a necessidade de ter um especialista que atue no sistema, o alto custo para manter um especialista na organização, e principalmente, a dificuldade de investir tempo e esforço na tarefa de descoberta de necessidades de aprendizagem dos desenvolvedores. Estas limitações deram origem a esta dissertação de mestrado que tem como objetivo auxiliar o especialista na tarefa de descoberta de assuntos para aprendizagem dos desenvolvedores e criação de esquemas de aprendizagem para o Ambiente Colaborativo Semântico voltado à empresas de desenvolvimento de software. Baseando-se em tecnologias da web 2.0 (motores de busca), mineração de textos e sistemas de recomendação, teorias e modelos de design instrucional, foram definidas abordagens para coleta de temas relevantes e para elaboração de cursos. Por fim, um ambiente foi implementado e dois estudos experimentais foram conduzidos, um estudo para avaliar a abordagem de coleta de informações e outro para avaliar a proposta de elaboração de cursos. Os resultados mostram que a abordagem é viável para a descoberta de assuntos para criação de cursos, especialmente em termos de elaboração de cursos organizacionais.

Palavras-chaves: Aprendizagem Organizacional, Ambiente Colaborativo Semântico, Motor de Busca, Mineração de Texto, Sistema de Recomendação, Engenharia de Software.

## ABSTRACT

Organizational learning is an area that helps companies to improve their processes significantly through the reuse of experiences, making knowledge accessible to the whole organization. However, to make learning possible in software development companies is not an easy task since it is an area in which processes and knowledge are usually internalized in the mind of their employees. The literature presents several works looking for ways to facilitate and encourage the sharing of information and the dissemination of knowledge. Among these works is the Semantic Collaborative Environment. This particular work is aimed to organizing the information generated by developers through a social tool (wiki), and thus, with the help of an expert able to implement learning schemes, create learning units that will be treated as courses, lessons or any limited part of a specific subject. However, the dependency of an expert on the learning schemes creation phase brought a number of limitations to this approach, for example, the necessity of having an expert acting on the system, the high cost position to maintain an expert on the organization, and mainly, the time spending and effort on discovering the developer learning needs. These limitations gave origin to this work, which aims to help the expert on the discovery of the developer learning needs, and the learning scheme creation task to the Semantic Collaborative Environment aimed to software development companies. Based on web 2.0 technologies (search engines), text mining and recommendation systems, models and theories of instructional design, approaches were defined in order to collect relevant subjects to the developer learning and the elaboration of courses. Lastly, an environment was implemented. Two experimental studies were conducted firstly to evaluate the approach of collecting subjects and after evaluate the elaboration courses proposal. The results show that the approach is viable for discovering subjects to the courses creation and it is efficient, especially in terms of elaborating organizational courses.

Key-words: Organizational Learning, Semantic Collaborative Environment, Search Engine, Text Mining, Recommendation System, Software Engineering.



## SUMÁRIO

<b>RESUMO</b> .....	<b>VII</b>
<b>ABSTRACT</b> .....	<b>VIII</b>
<b>LISTA DE FIGURAS</b> .....	<b>XII</b>
<b>LISTA DE QUADROS E TABELAS</b> .....	<b>XIV</b>
<b>LISTA DE ABREVIATURAS E SIGLAS</b> .....	<b>XV</b>
<b>CAPÍTULO 1 - INTRODUÇÃO</b> .....	<b>1</b>
1.1 - MOTIVAÇÃO.....	4
1.2 - OBJETIVOS.....	7
1.3 - DELIMITAÇÃO DO ESCOPO.....	8
1.4 - PROCESSOS DE TRABALHO .....	8
1.5 - ESTRUTURA DO DOCUMENTO DA DISSERTAÇÃO .....	9
1.6 - CONSIDERAÇÕES FINAIS .....	11
<b>CAPÍTULO 2 - APRENDIZAGEM ORGANIZACIONAL APLICADA A ENGENHARIA DE SOFTWARE</b> .....	<b>12</b>
2.1 - INTRODUÇÃO .....	12
2.2 - APRENDIZAGEM ORGANIZACIONAL.....	13
2.3 - AMBIENTES E MODELOS VOLTADOS À APRENDIZAGEM ORGANIZACIONAL.....	17
2.4 - AMBIENTE COLABORATIVO SEMÂNTICO.....	<b>ERRO! INDICADOR NÃO DEFINIDO.</b>
2.5 - REPRESENTAÇÃO DO CONHECIMENTO .....	18
2.5.1. <i>Unidades de Aprendizagem</i> .....	20
2.5.2. <i>Content Packages</i> .....	21
2.6 - ADAPTAÇÃO DE PADRÕES DE REPRESENTAÇÃO DE CONTEÚDOS PARA O AMBIENTE COLABORATIVO SEMÂNTICO.....	23
2.7 - O AMBIENTE COLABORATIVO SEMÂNTICO .....	26
2.7.1. <i>Pré-processamento</i> .....	27
2.7.2. <i>Gerador de Curso</i> .....	28
2.7.3. <i>Componente de Busca</i> .....	28
2.8 - A ARQUITETURA DO AMBIENTE .....	<b>ERRO! INDICADOR NÃO DEFINIDO.</b>
2.8.1. <i>Camada de Aplicação</i> .....	<b>Erro! Indicador não definido.</b>
2.8.2. <i>Camada de Memória Organizacional</i> .....	<b>Erro! Indicador não definido.</b>
2.9 - EDUCAÇÃO EM ORGANIZAÇÕES E DESIGN INSTRUCIONAL .....	29
2.10 - CONSIDERAÇÕES FINAIS .....	33
<b>CAPÍTULO 3 - DESCOBERTA DE CONHECIMENTO BASEADAS EM TEXTO</b> .....	<b>34</b>
3.1 - INTRODUÇÃO .....	34

3.2 - MOTORES DE BUSCA .....	36
3.3 - MINERAÇÃO DE TEXTO .....	40
3.3.1. <i>Descoberta de conhecimento a partir de texto</i> .....	41
3.3.2. <i>O algoritmo de clusterização CLIQUE</i> .....	43
3.3.3. <i>Medidas de Similaridade</i> .....	46
3.4 - SISTEMAS DE RECOMENDAÇÃO .....	47
3.5 - DESCOBERTA DE CONHECIMENTO E APRENDIZAGEM ORGANIZACIONAL .....	51
3.6 - DESCOBERTA DE CONHECIMENTO PARA ENGENHARIA DE SOFTWARE .....	52
3.7 - CONSIDERAÇÕES FINAIS .....	54
<b>CAPÍTULO 4 - ESTRUTURAÇÃO DA PESQUISA .....</b>	<b>55</b>
4.1 - ESTRUTURAÇÃO DA PESQUISA .....	55
4.1.1. <i>Planejamento Inicial</i> .....	56
4.1.2. <i>Fase Exploratória</i> .....	57
4.1.3. <i>Desenvolvimento</i> .....	59
4.1.4. <i>Avaliação e Análise</i> .....	60
4.2 - CONSIDERAÇÕES FINAIS .....	65
<b>CAPÍTULO 5 - O AMBIENTE GERADOR DE ESQUEMA DE APRENDIZAGEM .....</b>	<b>66</b>
5.1 - INTRODUÇÃO .....	66
5.2 - IDENTIFICADOR DE TEMAS E PAPÉIS .....	69
5.2.1. <i>Arquitetura do Mecanismo Identificador de Temas</i> .....	78
5.3 - DEFINIDOR DE CURSO .....	79
5.3.1. <i>Criação do Esquema de Aprendizagem</i> .....	84
5.4 - CONSIDERAÇÕES FINAIS .....	89
<b>CAPÍTULO 6 - EXPERIMENTO PARA AVALIAR A IDENTIFICAÇÃO DE TEMAS .....</b>	<b>90</b>
6.1 - INTRODUÇÃO .....	90
6.2 - O EXPERIMENTO .....	90
6.3 - O MÉTODO EXPERIMENTAL .....	91
6.4 - DEFINIÇÃO DOS OBJETIVOS .....	92
6.5 - PLANEJAMENTO .....	92
6.5.1. <i>Definição das hipóteses</i> .....	93
6.5.2. SELEÇÃO DOS PARTICIPANTES .....	93
6.5.3. DESCRIÇÃO DO EXPERIMENTO .....	94
6.5.4. DESCRIÇÃO DA ANÁLISE .....	95
6.5.5. <i>Execução</i> .....	99
6.6 - RESULTADOS .....	99
6.6.1. <i>Mecanismo de Sugestão de Temas</i> .....	100
6.6.2. <i>Unidades de Análise</i> .....	103
6.7 - Discussão .....	105

6.8 - CONSIDERAÇÕES FINAIS .....	107
<b>CAPÍTULO 7 - EXPERIMENTO PARA AVALIAÇÃO DO AMBIENTE.....</b>	<b>108</b>
7.1 - INTRODUÇÃO .....	108
7.2 - O EXPERIMENTO .....	109
7.3 - O MÉTODO EXPERIMENTAL .....	109
7.4 - DESCRIÇÃO DOS OBJETIVOS .....	109
7.5 - PLANEJAMENTO .....	110
7.5.1. <i>Definição das Hipóteses</i> .....	110
7.5.2. <i>Seleção dos Participantes</i> .....	110
7.5.3. <i>Descrição do Experimento</i> .....	111
7.5.4. <i>Execução</i> .....	113
7.6 - RESULTADOS.....	114
7.6.1. <i>Definição de Cursos</i> .....	114
7.7 - DISCUSSÃO .....	119
7.8 - CONSIDERAÇÕES FINAIS .....	119
<b>CAPÍTULO 8 - CONCLUSÕES.....</b>	<b>121</b>
<b>8.1 - RELEVÂNCIA DO ESTUDO .....</b>	<b>121</b>
8.2 - CONTRIBUIÇÕES DA PESQUISA.....	121
8.3 - LIMITAÇÕES.....	122
8.4 - TRABALHOS FUTUROS.....	123
<b>REFERÊNCIAS BIBLIOGRÁFICAS.....</b>	<b>124</b>
<b>APÊNDICE A – QUESTIONÁRIO APLICADO NO EXPERIMENTO DE AVALIAÇÃO DA ABORDAGEM DE IDENTIFICAÇÃO DE TEMAS E PAPÉIS.....</b>	<b>143</b>
<b>APÊNDICE B – PROJETO DE SISTEMA FICTÍCIO – PROCEDIMENTO DE EXPERIMENTO APLICADO NO EXPERIMENTO DE AVALIAÇÃO DA ABORDAGEM DE CRIAÇÃO DE CURSOS .....</b>	<b>145</b>

## LISTA DE FIGURAS

Figura 1-1. Etapas de Geração de Unidades de Aprendizagem. Adaptado de (MENOLLI; REINEHR; MALUCELLI, 2013b) .....	5
Figura 2-1. Processo de conversão do conhecimento, adaptado de (NONAKA; TAKEUCHI, 1997) .....	14
Figura 2-2. Modelo conceitual do IMS-LD (IMS, 2003) .....	21
Figura 2-3. Escopo do IMS Content Packaging. (IMS, 2004) .....	22
Figura 2-4. Diagrama de Classes do Learning Design da UOLO. Adaptado de (MENOLLI et al., 2014). .....	25
Figura 2-5. Diagrama de Classes do Content Package da UOLO. ....	25
Figura 2-6. Esquema para Geração de Unidades de Aprendizagem Organizacionais (MENOLLI, 2013).....	26
Figura 2-7. Esquema de Ambiente Colaborativo Semântico (MENOLLI, 2013). .....	27
Figura 2-8. Exemplo de Esquema de Aprendizagem definido pelo Especialista (MENOLLI; MALUCELLI; REINEHR, 2011). .....	<b>Erro! Indicador não definido.</b>
Figura 2-9. Busca por Objetos e Unidades de Aprendizagem Organizacionais. Adaptado de (MENOLLI, 2013). .....	<b>Erro! Indicador não definido.</b>
Figura 2-10. Arquitetura do Ambiente Colaborativo Semântico. Adaptado de (MENOLLI; MALUCELLI; REINEHR, 2011). .....	<b>Erro! Indicador não definido.</b>
Figura 3-1. Grafo do Algoritmo Cliques (KOWALSKI, 2007). .....	46
Figura 3-2. Processo de Recomendação Baseada em Conteúdo (LOPS; GEMMIS; SEMERARO, 2011). .....	49
Figura 3-3. Processo de Recomendação Baseada no Usuário (LESKOVEC; RAJARAMAN; ULLMAN, 2014).....	50
Figura 4-1. Estrutura da Pesquisa.....	56
Figura 5-1. Gerador de Cursos (MENOLLI; REINEHR; MALUCELLI, 2013b)..	67
Figura 5-2. Extensão do Gerador do Esquema de Aprendizagem. ....	68
Figura 5-3. Crescimento de Acessos ao Fórum Online StackOverflow (QuantaCast, 2015).....	70
Figura 5-4. Servidor Proxy com SQUID. ....	72
Figura 5-5. Amostra de registro do arquivo de <i>logs</i> de acesso ao Google. ....	73

Figura 5-6. Arquitetura do Mecanismo Identificador de Temas.....	78
Figura 5-7. Tela para filtro de temas sugeridos por período.....	80
Figura 5-8. Tela de Cadastro de Habilidades de Desenvolvedores. ....	82
Figura 5-9. Etapas para Criação do Curso. Passo 1 - Informações Básicas....	85
Figura 5-10. Etapas de Criação do Curso. Passo 2 - Temas .....	86
Figura 5-11. Etapas de Criação do Curso. Passo 3 - Organização.....	87
Figura 5-12. Estrutura dos temas. ....	87
Figura 5-13. Etapas de Criação do Curso. Passo 4 - Papéis .....	88
Figura 6-1. Estrutura Geral da Análise Discursiva Textual.....	98

## LISTA DE QUADROS E TABELAS

Tabela 2-1. Comparativo das Propriedades do UOLO e do IMS LD. Adaptado de (MENOLLI et al., 2014).....	<b>Erro! Indicador não definido.</b>
Tabela 2-2. Comparativo das propriedades UOLO e IMS CP. Adaptado de (MENOLLI, 2013). .....	<b>Erro! Indicador não definido.</b>
Tabela 6-1. Exemplos de palavras antes e após o processo de Stemming. ....	96
Quadro 2-1. Trabalhos voltados à Aprendizagem Organizacional (MENOLLI, 2013). .....	<b>Erro! Indicador não definido.</b>
Quadro 2-2. Características de um Ambiente de Aprendizagem Organizacional. Adaptado de (MENOLLI, MALUCELLI, REINEHR, 2011)	<b>Erro! Indicador não definido.</b>
Quadro 2-3. Manifesto do Padrão IMS CP. Adaptado de (COSTA et al., 2008) .....	22
Quadro 3-1. Motores de Busca mais utilizados no mundo. ....	38
Quadro 3-2. Métodos mais utilizados para clusterização de textos (AGGARWAL; ZHAI, 2012). .....	44
Quadro 4-1. Termos da Pesquisa (MENOLL; REINEHR; MALUCELLI, 2013a). .....	58
Quadro 4-2. Métodos de Validação para Experimentação. Adaptado de (ZELKOWITZ; WALLACE; BINKLEY, 2003) .....	63
Quadro 5-1. URLs originadas de pesquisas realizadas no motor de busca Google.....	74
Quadro 6-1. Questionário aplicado no experimento para avaliação da abordagem. ....	97

## LISTA DE ABREVIATURAS E SIGLAS

ADS	Ambiente de Desenvolvimento de Software
AO	Aprendizagem Organizacional
API	<i>Application Programming Interface</i>
CBR	Raciocínio Baseado em Casos
CIO	<i>Chief of Information Office</i>
CRT	Coletor de Requisitos Técnicos
DC	Definidor de Cursos
FTP	<i>File Transfer Protocol</i>
HTML	<i>Hypertext Markup Language</i>
HTTP	<i>Hypertext Transfer Protocol</i>
HTTPS	<i>Hypertext Transfer Protocol Secure</i>
IEEE	<i>Institute of Electrical and Eletronics Engineers</i>
IP	<i>Internet Protocol</i>
ISO	<i>International Organization for Standardization</i>
ITP	Identificador de Temas e Papéis
KDD	<i>Knowledge Discovery Databases</i>
KDT	<i>Knowledge Discovery from Text</i>
LD	<i>Learning Design</i>
LME	Linguagens de Modelagem Educacional
LMML	<i>Learning Material Markup Language</i>
LOM	<i>Learning Objects Metadata</i>
LTSC	<i>Learning Technology Standards Committee</i>
MIC	<i>Meta model of Instructional Context</i>
OAs	Objetos de Aprendizagem
OLO	<i>Organizational Learning Object</i>
OOLO	<i>Ontology for Organizational Learning Object</i>
PLN	Processamento de Linguagem Natural
RDF	<i>Resource Description Framework</i>
SSEW	<i>Semantic Software Engineering Wiki</i>
TCP/IP	<i>Transmission Control Protocol / Internet Protocol</i>
TI	Tecnologia de Informação
UOLO	<i>Unit of Organizational Learning Ontology</i>

URL	<i>Uniform Resource Locator</i>
W3C	<i>World Wide Web Consortium</i>
XML	<i>Extensible Markup Language</i>



## Capítulo 1 - INTRODUÇÃO

Gerentes de Tecnologia da Informação (TI) e *Chief of Information Office* (CIOs) têm, cada vez mais, que justificar os gastos no orçamento e demonstrar os benefícios produzidos e trazidos para a companhia (CIO Association, 2014). Resultados de uma pesquisa indicam que nos últimos quatro anos, empresas de desenvolvimento de software adotaram abordagens cuidadosas com foco em melhoramentos em produtividade e habilidades, assim como em eficiência de processos internos e agilidade (CIO Association, 2014). O melhoramento de produtividade e melhor capacitação dos colaboradores podem ser realizados a partir de treinamentos internos e gerenciamento do conhecimento adquirido ao longo do tempo.

Por este motivo, é importante considerar o treinamento de talentos e aprendizado na organização não apenas como um investimento, mas também como algo que possa significar economia dos custos em médio prazo. A capacitação dos colaboradores, aliada ao desenvolvimento de habilidades em prol do gerenciamento do conhecimento, reflete diretamente na evolução de organizações perante o mercado.

O impacto do conhecimento nas organizações é tão relevante que é tratado não mais como um fator estratégico em potencial, acessível a poucos privilegiados, mas como um elemento comum, porém essencial à sobrevivência organizacional (NEVES, 2011). Isso faz com que o conhecimento seja de vital importância para corporações e o produto seja gerado a partir do conhecimento intensivo da empresa. Os projetos de conhecimento intensivo referem-se aqueles em que a maioria dos trabalhos é dito ser de natureza intelectual e funcionários qualificados formam a maior parte da força de trabalho (ALVESSON, 2000).

No desenvolvimento de software, o domínio técnico de cada empregado, juntamente com o conhecimento que estes possuem das práticas e rotinas empresariais, faz com que estes profissionais tenham um grande

valor para a organização. Assim, com o tempo, os profissionais de software tornam-se mais valorizados pelas experiências adquiridas e lições aprendidas, fazendo com que se tornem uma fonte de conhecimento fundamental para a empresa. Contudo, o alto valor dado a esses empregados, faz com que outras empresas tenham interesse no conhecimento desse profissional, o que causa grande rotatividade no mercado de trabalho. Perder um funcionário experiente e qualificado para outra empresa significa perder o conhecimento adquirido ao longo do tempo (MENOLLI; MALUCELLI; REINEHR, 2011).

Esta realidade faz com que as organizações busquem meios para armazenar este conhecimento e compartilhar a base de conhecimento para a organização como um todo. A área que busca minimizar estes problemas é a Aprendizagem Organizacional (AO), que trata da capacidade ou dos processos dentro da organização, destinados a manter ou melhorar o desempenho com base na experiência (NEVIS; DI BELLA; GOULD, 1995).

Vários trabalhos na área de aprendizagem organizacional foram propostos no decorrer dos anos, principalmente a partir dos anos 90, os quais tentam encontrar fatores que auxiliem este processo. No contexto da Engenharia de Software há vários estudos que abordam este tema. Um dos estudos avalia o desempenho de ontologias para a classificação de materiais de aprendizagem de acordo com o estilo de aprendizagem e a área da Engenharia de Software (VALASKI; MALUCELLI; REINEHR; SANTOS, 2011); outra pesquisa auxilia na comunicação entre as equipes de desenvolvimento de software (KRONE; SYVÄJÄRVI; STENVALL, 2009); há também trabalhos que contribuem para o gerenciamento do conhecimento nas organizações de software, como (SHAHZAD, 2009); e alguns estão relacionados com ambientes de *e-learning* utilizando recursos semânticos, como por exemplo (CAPUANO; MIRANDA; ORCIUOLI, 2009).

Recentemente a utilização de tecnologias Web também tem sido uma opção, principalmente em empresas de software, pois permitem a criação de ferramentas sociais como redes sociais, *wikis* e *blogs*. Por meio destas ferramentas, em especial as de edição colaborativa, é possível que qualquer pessoa possa contribuir para a criação de conhecimento, e assim este conhecimento é compartilhado e acessado, se tornando informação útil para a comunidade (MENOLLI; MALUCELLI; REINEHR, 2011).

No entanto, o uso das ferramentas colaborativas apenas contribui para prover um ambiente social colaborativo, não sendo suficiente para atingir as características necessárias para que a aprendizagem organizacional ocorra (MENOLLI; MALUCELLI; REINEHR, 2011).

Sendo assim, além das novas tecnologias e ferramentas baseadas na Web, é necessário organizar as informações de maneira que possam auxiliar o aprendizado. No trabalho de Sicilia e Lyras (SICILIA; LYTRAS, 2005), é descrito que do ponto de vista de uma empresa, a aplicação de tecnologias de Web Semântica devem ser motivadas e melhoradas para mecanismos orientados ao aprendizado, incluindo aspectos culturais. Outros conceitos que podem ser empregados para melhorar a aprendizagem são conceitos advindos da área educacional, mais especificamente do ensino à distância, tais como, ambientes de *e-learning* que são ambientes de ensino não presencial suportado por tecnologia, objetos de aprendizagem e unidades de aprendizagem (POLSANI, 2004).

Os Objetos de Aprendizagem (OAs) são definidos como qualquer tipo de recurso, digital ou não, que possui o intuito de auxiliar o processo de ensino e aprendizagem, podendo ser reutilizado em diferentes contextos (IEEE, 2002). A partir de Objetos de Aprendizagem é possível definir as Unidades de Aprendizagem, que é um termo abstrato usado para se referir a qualquer parte delimitada de uma formação, como um curso, um módulo ou uma lição (BRASIL, 2007). Para Koper e Olivier (KOPER; OLIVIER, 2004), as Unidades de Aprendizagem podem ser reutilizadas por pessoas ou grupos diferentes e em situações variadas, principalmente em um ambiente *online*.

Baseado neste contexto, o trabalho de Menolli, Reinehr e Malucelli (MENOLLI; REINEHR; MALUCELLI, 2013b) propõe um ambiente colaborativo semântico para empresas de desenvolvimento de software. Este ambiente tem como principal objetivo organizar o conhecimento que está armazenado de forma textual, e a partir deste conhecimento auxiliar na aprendizagem dos indivíduos da organização (MENOLLI, 2013). A proposta de auxílio na aprendizagem organizacional se dá por meio da utilização de objetos de aprendizagem, definidos a partir de documentos de *wiki* para a geração de cursos, por meio da definição de unidades de aprendizagem. Porém, a geração

de cursos depende da disponibilidade de um especialista na área de domínio para definir a estrutura de um Esquema de Aprendizagem.

Um Esquema de Aprendizagem é uma estrutura definida em uma metalinguagem, por exemplo, XML (*Extensible Markup Language*). O XML é uma linguagem de marcação tal qual o HTML (*Hypertext Markup Language*), portanto possui *tags* que são utilizadas para descrição de dados (W3 SCHOOLS, 2015b). As *tags* definidas para o Esquema de Aprendizagem contêm elementos que formam um processo de ensino e aprendizagem (IMS, 2003). Estes elementos formadores de processo de ensino são, por exemplo, informações sobre o curso, tais como o objetivo e pré-requisito, além das atividades de aprendizagem e sua hierarquia e sequência de realização (IMS, 2003).

Entretanto, mesmo que a empresa possua um especialista para a definição do Esquema de Aprendizagem, isso não garante que este especialista tenha conhecimento da real necessidade de aprendizagem da equipe. Sendo assim, faz-se necessário mapear as necessidades da equipe, o que exige tempo e esforço, podendo ser uma barreira para a criação de um Esquema de Aprendizagem, e conseqüentemente para o uso do ambiente colaborativo.

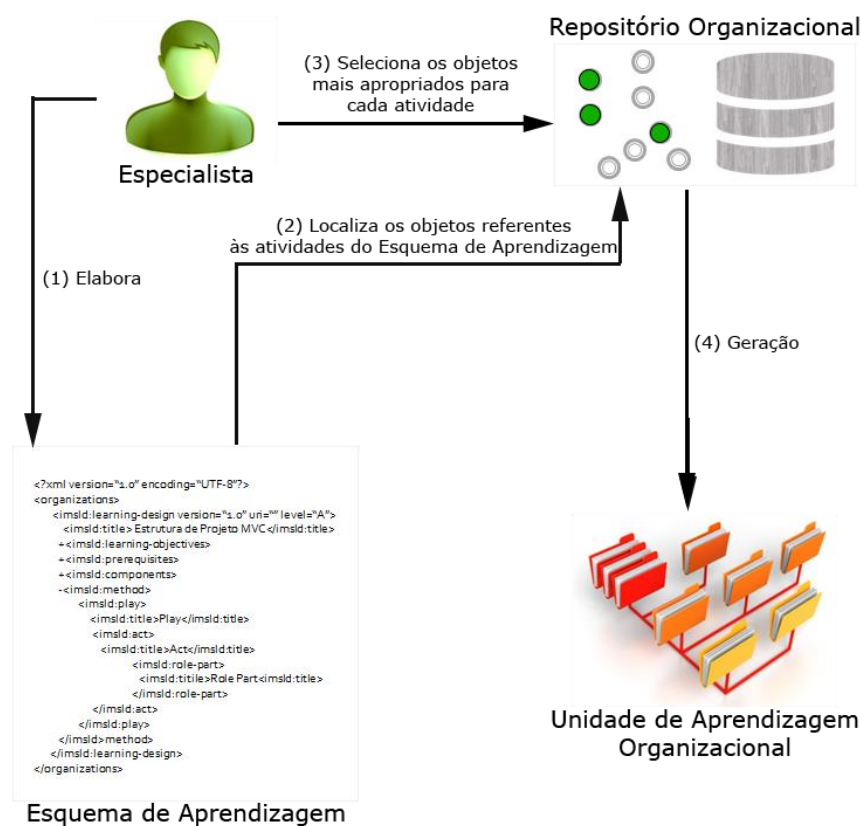
Outra dificuldade identificada para o uso do ambiente é a complexidade em elaborar um Esquema de Aprendizagem, visto que o esquema é formado por estruturas em uma metalinguagem. Formular o Esquema de Aprendizagem, garantindo que todas as suas estruturas sejam respeitadas em sua sintática e semântica, torna-se trabalhoso, o que pode no dia a dia, se tornar uma tarefa que o especialista opte em não executar.

Portanto, tentando minimizar estas dificuldades, esta dissertação apresenta uma proposta para que a tarefa de elaboração do Esquema de Aprendizagem por parte do especialista seja feita de maneira dinâmica,

## **1.1 - Motivação**

No ambiente proposto por Menolli, Reinehr e Malucelli (MENOLLI; REINEHR; MALUCELLI, 2013b), para que seja definida uma Unidade de

Aprendizagem, o componente Gerador de Unidade de Aprendizagem lê um Esquema de Aprendizagem que deve ser inicialmente criado por um especialista. Para cada atividade descrita de forma estruturada no Esquema de Aprendizagem, o ambiente retorna uma lista de Objetos de Aprendizagem correspondentes. Ao fim desta atividade o especialista terá uma Unidade de Aprendizagem, que agrupa os Objetos de Aprendizagem selecionados. Esta Unidade poderá ser encontrada na busca do ambiente proposto em Menolli, Reinehr e Malucelli (MENOLLI; REINEHR; MALUCELLI, 2013b). A Figura 1-1 descreve detalhadamente os passos para gerar uma unidade de aprendizagem.



**Figura 1-1. Etapas de Geração de Unidades de Aprendizagem. Adaptado de (MENOLLI; REINEHR; MALUCELLI, 2013b)**

Nota-se neste ambiente, um envolvimento importante por parte do especialista, que tem o papel de criar o Esquema de Aprendizagem. A execução correta desta tarefa é crucial para que o ambiente seja capaz de

encontrar, no repositório da organização, os Objetos de Aprendizagem correspondentes.

No entanto, a alta complexidade exigida para gerar um Esquema de Aprendizagem (representado em XML), que tenha em sua estrutura as definições de hierarquia, temas, requisitos, entre outros componentes, se torna uma barreira na tarefa de geração de Unidades de Aprendizagem.

Outro obstáculo ligado à presença de um especialista na organização é o esforço envolvido em investigar a real necessidade de aprendizagem dos colaboradores em determinado momento. Este obstáculo está relacionado com um amplo campo de estudo, que são as Organizações que Aprendem. Garvin (GARVIN, 1993), assumindo um ponto de vista cognitivo/comportamental, define “Organizações de Aprendizagem” (*learning organizations*) como aquelas capazes de adquirir, criar, produzir, transferir e modificar conhecimentos e comportamentos de seus membros. Shein (SHEIN, 1996) acrescenta ao rol de elementos que criam obstáculos à efetivação de Organizações de Aprendizagem, o aspecto da comunicação. Para ele, as organizações falham em aprender principalmente pela incapacidade de criar canais de comunicação que permitam a negociação de interesses entre o colaborador e o gestor.

Além das limitações citadas, a dependência de um especialista na organização apresenta outras barreiras, como o alto custo de um especialista na organização e a alta carga de trabalho que é atribuída ao especialista no dia a dia na atuação na organização. Essa atribuição prejudica a antecipação das ações, principalmente quando a equipe precisa de treinamento para atuar em um projeto futuro.

Sendo assim, é importante estudar e propor uma forma de potencializar o Ambiente Colaborativo Semântico proposto em Menolli, Malucelli e Reinehr (MENOLLI; MALUCELLI; REINEHR, 2011), para que seja possível colaborar ainda mais com a aprendizagem organizacional. Quando um usuário tem uma dificuldade para solucionar um problema, geralmente ele realiza pesquisas na internet por materiais que possam auxiliá-lo. Seria interessante capturar as pesquisas realizadas e sugerir ao especialista os temas mais procurados e que precisam de um treinamento.

Também é observada a necessidade de criar uma abordagem que auxilie o especialista na tarefa de criação do esquema de aprendizagem a

partir dos temas considerados mais relevantes. Sendo assim, é possível contribuir para a aprendizagem organizacional, construindo as habilidades dos desenvolvedores e conseqüentemente, criando o esquema de aprendizagem que será o objeto de integração entre o ambiente proposto neste trabalho e o ambiente colaborativo semântico proposto em Menolli; Malucelli e Reinehr (MENOLLI; MALUCELLI; REINEHR, 2011).

## **1.2 - Objetivos**

Em (MENOLLI; REINEHR; MALUCELLI, 2013a), são identificados por meio de uma revisão sistemática, vários trabalhos de engenharia de software que utilizam os conceitos de aprendizagem organizacional. Dentre os estudos identificados, pode-se verificar que há uma prevalência em estudos direcionados a processos de software, bem como pesquisas com foco em desenvolvimento distribuído. Também há estudos que apresentam o uso de ferramentas semânticas, a fim de aprimorar o aprendizado em projetos de engenharia de software, e pesquisas que apresentam o uso de conceitos da área educacional, porém, sem serem aplicados para a contribuição da aprendizagem organizacional.

Portanto, há a necessidade em trabalhos de aprendizagem organizacional, voltados à engenharia de software, que contribuam para potencializar os ambientes de aprendizagem organizacional no sentido de melhorar os processos de compartilhamento de conhecimento, principalmente quando se trata de propostas em que, sabendo da importância que o especialista tem para a aprendizagem organizacional, o foco é encontrar formas de tornar o papel desempenhado pelo especialista mais automatizado.

Estes fatores contribuem para a definição do objetivo geral deste trabalho que é desenvolver um mecanismo para criação semiautomática do esquema de aprendizagem, em um ambiente colaborativo semântico, voltado à aprendizagem organizacional para empresas de desenvolvimento de software.

Para atender ao objetivo principal da pesquisa, são apresentados os seguintes objetivos específicos a serem atingidos:

- I. Desenvolver um mecanismo para captar as consultas em motores de busca.
- II. Classificar as consultas realizadas pelos desenvolvedores em temas relevantes para geração de cursos.
- III. Criar semiautomaticamente o esquema de aprendizagem.
- IV. Avaliar o mecanismo proposto.

### **1.3 - Delimitação do Escopo**

Inicialmente é necessário definir o universo de empresas que é objeto da proposta. O trabalho proposto por (MENOLLI; MALUCELLI; REINEHR, 2011), tem como escopo empresarial apenas organizações desenvolvedoras de software, sendo assim, por se tratar de uma ampliação deste ambiente já proposto, o desígnio deste trabalho se mantém voltado para organizações nas quais a atividade desenvolvida pelos seus funcionários envolve a engenharia de software.

Igualmente se faz necessário restringir o propósito deste trabalho de forma a limitar seu escopo, uma vez que a aprendizagem organizacional pode ser atingida de diferentes formas, algumas sem a ajuda de tecnologia. Este trabalho não se destina a descrever a melhor maneira de se alcançar a aprendizagem organizacional, ou ser a única maneira de fazê-lo, já que diversos meios estimulam de diferentes formas a aprendizagem organizacional.

Por este trabalho ser uma proposta técnica para um problema que tem sua origem, em alguns casos, em aspectos culturais, este trabalho não tem como objetivo sugerir ou alterar políticas empresariais, formas de interação interpessoal e valores organizacionais.

Por fim, não faz parte do escopo deste trabalho a implantação deste estudo diretamente em um ambiente empresarial, haja vista, que para tanto, existem fatores sociais, culturais, psicológicos e organizacionais que influenciam esta análise.

### **1.4 - Processos de Trabalho**



De maneira a organizar o trabalho de pesquisa a ser realizado, foi definido um processo inicial contendo fases necessárias para atingir os objetivos propostos, que são:

- Fase 1 – Preparação da pesquisa: fase que corresponde à delimitação da área de estudo, coleta e análise das referências bibliográficas com objetivo de determinar o foco do trabalho e os objetivos a serem atingidos.
- Fase 2 – Estruturação da pesquisa: fase de elaboração de um quadro referencial teórico, seleção do método de pesquisa e definição das etapas de pesquisa.
- Fase 3 – Execução da pesquisa: fase da investigação em si, com procura e análise de trabalhos na literatura, bem como a proposta de uma abordagem para criar a identificação do conhecimento específico dos funcionários de uma organização, assim como a implementação da abordagem proposta e o mecanismo de sugestão para criação de conteúdos propostos.
- Fase 4 – Avaliação e Análise dos Resultados: fase da realização da avaliação do mecanismo proposto e análise dos resultados extraindo as limitações da pesquisa, conclusões e trabalhos futuros.

## **1.5 - Estrutura do documento da dissertação**

O Capítulo 1, aqui apresentado, visa oferecer ao leitor um panorama geral sobre o contexto no qual se insere este trabalho de pesquisa. Visa, ainda, estabelecer o objetivo geral e os objetivos específicos, bem como apresentar os processos do trabalho.

O Capítulo 2 aprofunda o referencial teórico inicial descrito no Capítulo 1, sobre aprendizagem organizacional e como é utilizada na engenharia de software. Este capítulo também apresenta de forma conceitual e técnica, o ambiente colaborativo social semântico proposto em (MENOLLI; MALUCELLI; REINEHR, 2011) que é aplicado como base para o desenvolvimento deste estudo.

No Capítulo 3 são introduzidos alguns princípios e as principais teorias sobre tecnologias e técnicas relacionadas à Descoberta de Conhecimento, como Recuperação de Informação, Sistemas de Recomendação, Mineração de Texto e Motores de Busca. Estes conhecimentos são necessários, pois servem de base para o desenvolvimento das abordagens propostas neste trabalho.

O Capítulo 4 apresenta alguns conceitos importantes referentes à metodologia e métodos de pesquisa, de forma a posicionar o leitor quanto às opções metodológicas adotadas e detalhar as estratégias de pesquisa para atingir o objetivo geral do trabalho.

O Capítulo 5 apresenta a arquitetura do mecanismo de coleta dos temas e papéis, assim como descreve todos os seus componentes. Além disso, apresenta também a abordagem utilizada para gerar o esquema de aprendizagem semiautomaticamente para o ambiente colaborativo semântico voltado para empresas de desenvolvimento de software, bem como a implementação de todos os componentes.

O Capítulo 6 apresenta um experimento realizado para verificar a viabilidade de implantar um ambiente baseado na abordagem proposta, o qual coleta pesquisas realizadas por um grupo de desenvolvedores de uma empresa em um motor de busca, e os apresenta em forma de temas para aprendizado da equipe.

O capítulo 7 apresenta o experimento para avaliar o mecanismo implementado. A partir dos dados coletados neste experimento, em conjunto com os dados do experimento piloto, os objetivos de pesquisa propostos são discutidos.

O Capítulo 8 conclui este trabalho, apresentando as considerações finais, descrevendo a relevância do estudo e as contribuições da pesquisa. Este capítulo também expõe trabalhos futuros de pesquisa na área de aprendizagem organizacional em engenharia de software, em especial relacionadas ao ambiente proposto.

Por fim, os Apêndices contidos neste documento apresentam os documentos e questionários utilizados nos experimentos para avaliação das abordagens propostas e do ambiente.

## **1.6 - Considerações Finais**

Este capítulo apresentou a relevância da aprendizagem organizacional na engenharia de software e os motivos que levam as empresas a instaurar métodos que auxiliem na aprendizagem. Foram apresentadas algumas soluções, bem como o ambiente que está sendo utilizado como base para continuação da pesquisa na área de aprendizagem organizacional. Por fim, foram apresentadas as justificativas para o desenvolvimento deste estudo, sua problematização e objetivos.

## Capítulo 2 - **APRENDIZAGEM ORGANIZACIONAL APLICADA A ENGENHARIA DE SOFTWARE**

Neste capítulo é apresentada uma revisão da literatura sobre aprendizagem organizacional, mais precisamente aprendizagem organizacional na engenharia de software. Também é exposto o Ambiente Colaborativo Social Semântico (MENOLLI, 2013) que serve como base para alcançar o objetivo deste trabalho. Este capítulo, também apresenta um paradigma sobre a educação em organizações, bem como alguns modelos de design instrucional que foram aplicados neste trabalho.

### **2.1 - Introdução**

Os projetos de software são capazes de oferecer aos seus participantes novas experiências e novas lições, desde que a tarefa exija a utilização de uma nova ferramenta, um novo paradigma de programação ou até mesmo uma evolução dos processos dessa organização (TACLA; BARTHÈS, 2003). Estas experiências adquiridas ao longo do tempo em uma organização fazem com que o conhecimento deste funcionário seja de grande valor e também uma propriedade fundamental para as empresas na economia contemporânea. Cada vez mais o conhecimento é distribuído entre os indivíduos, equipes e organizações. Conseqüentemente, a aptidão para criar, conquistar, integrar, implantar e difundir o conhecimento tem surgido como uma capacidade organizacional fundamental (TAKEISHI, 2002) (TEECE; PISANO; SHUEN, 1997).

Porém, existem muitas barreiras para uma implantação bem sucedida de ferramentas que contribuam para o armazenamento de conhecimento. Tipos de documentos podem ser definidos, bancos podem ser integrados, ontologias acordadas, ferramentas colaborativas aplicadas, mas, sem integrar tudo isso

na vida organizacional dos indivíduos interessados, estas ferramentas permanecerão sem uso e com o tempo serão esquecidas.

Para a engenharia de software, por ser uma área de processos baseados no conhecimento, é de suma importância manter os processos de armazenamento e reutilização do conhecimento, pois a única vantagem sustentável que uma empresa tem é aquilo que ela coletivamente sabe (DAVENPORT; PRUSAK, 2003). Portanto, para manter-se no mercado com qualidade é necessário que as empresas voltem sua atenção para este aspecto.

Com isso, muitos trabalhos são aplicados na identificação de elementos que possam contribuir e até mesmo automatizar a aprendizagem em um ambiente corporativo para situações específicas.

Dessa forma, é aceito que os sistemas de informação podem auxiliar a aprendizagem organizacional, mas é uma área que ainda necessita de pesquisas, principalmente para problemas específicos.

## **2.2 - Aprendizagem Organizacional**

O principal objetivo em gerenciar o conhecimento, especificamente para a engenharia de software, é que este recurso possa auxiliar os usuários em tarefas que são de conhecimento intensivo. Para organizar o conhecimento é necessário que haja meios de facilitar a criação, estruturação, compartilhamento, distribuição e uso do aprendizado.

Para Nonaka e Takeuchi (NONAKA; TAKEUCHI, 1997), o conhecimento é o principal ativo no processo de aprendizagem organizacional, e pode ser classificado em dois tipos: o tácito e o explícito. O conhecimento tácito está incorporado à experiência individual e envolve fatores intangíveis como crenças pessoais, perspectivas e sistemas de valor. Já o conhecimento explícito pode ser associado em uma linguagem formal, como afirmações gramaticais, expressões matemáticas, especificações, manuais e outros, e a sua transmissão ocorre de maneira mais fácil entre as pessoas.

A transformação deste conhecimento, conforme afirmado por Nonaka e Takeuchi (NONAKA; TAKEUCHI, 1997), acontece de forma interativa e em espiral, assim como apresentado na Figura 2-1.

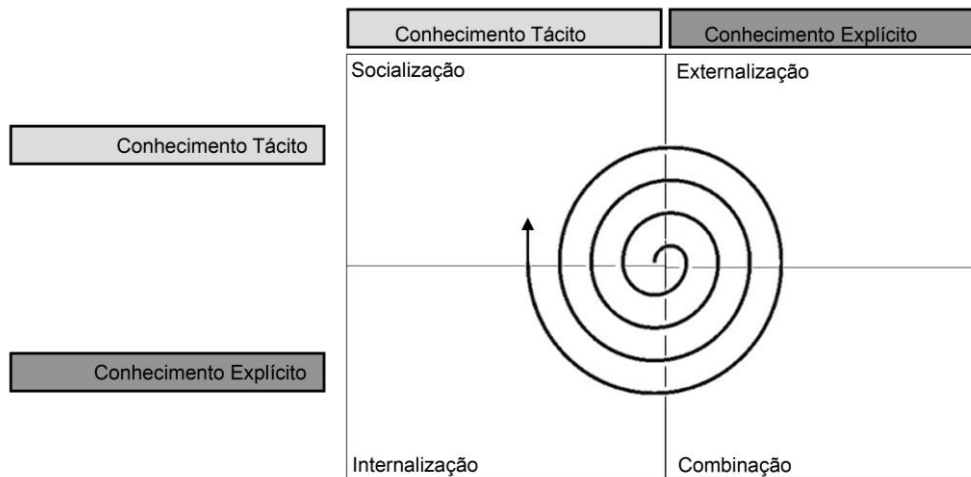


Figura 2-1. Processo de conversão do conhecimento, adaptado de (NONAKA; TAKEUCHI, 1997)

Neste modelo, é proposto que o processo de conversão de conhecimento é feito de quatro formas, que são geradas pela interação do conhecimento tácito e explícito. Cada uma das quatro formas de conversão é descrita a seguir (NONAKA; KONNO, 1998):

- socialização (tácito em tácito): neste processo há troca de experiências, modelos e técnicas mentais e habilidades. O indivíduo pode adquirir conhecimento sem uma linguagem formal, por esta ocorrer por meio de observação e prática;
- externalização (tácito em explícito): constitui o mais relevante processo de conversão de conhecimento e é expressado na forma de metáforas, analogias, conceitos, hipóteses ou modelos. Ele é acarretado pelo diálogo ou pela reflexão de um grupo. A externalização é a chave para criação do conhecimento, pois cria conceitos novos e explícitos a partir do conhecimento tácito;
- combinação (explícito em explícito): neste processo há a sistematização dos conceitos e a combinação de conjuntos diferentes de conhecimento. O conhecimento é compartilhado por meio de documentos, reuniões, conversas ao telefone ou redes de comunicação computadorizadas e a criação do conhecimento é realizada por meio de treinamento formal; e,
- internalização (explícito em tácito): neste processo há a inserção do conhecimento explícito ao conhecimento tácito, que se relaciona diretamente com o “aprender fazendo”, como por exemplo, quando

um técnico lê um manual e absorve o conhecimento a ponto de não precisar mais dele para realizar o trabalho.

Segundo (DAVENPORT; PRUSAK, 1998) as organizações passaram a valorizar cada vez mais a experiência e o *know-how* de seus funcionários. Este conhecimento implícito é aplicado de diversas formas nas empresas, como nas rotinas, práticas de produção e nos relacionamentos. Com isso, surgiu a necessidade de criar e implantar processos que gerem, armazenem, organizem, disseminem e apliquem o conhecimento produzido e utilizado na empresa de modo sistemático, explícito, confiável e acessível à comunidade da organização (MENOLLI, 2013).

Para que seja possível aplicar a implantação de processos para gerenciamento do conhecimento produzido pela organização, faz-se necessário o entendimento do conceito de aprendizagem organizacional. Aprendizagem organizacional é uma área de conhecimento dentro da teoria organizacional que estuda modelos e teorias sobre a forma como as organizações aprendem e se adaptam (VASENSKA, 2013).

Para Nevis, Di Bella e Gould (NEVIS; DI BELLA; GOULD, 1995), a aprendizagem organizacional, é a capacidade, ou os processos dentro da organização, destinados a manter ou melhorar o desempenho com base na experiência.

A aprendizagem organizacional tem uma importância fundamental na área de engenharia de software, uma vez que as empresas estão criando e aproveitando o conhecimento, dados e informações em um ritmo sem precedentes, um fenômeno que faz com que o uso do conhecimento e da tecnologia não seja uma opção, mas uma necessidade (DRUCKER, 1993).

No processo de desenvolvimento de software, a aprendizagem organizacional junto à gestão do conhecimento pode ser utilizada para apoiar melhor as atividades diversas, tais como definição de processo de software, alocação de recursos humanos, estimativa, análise de requisitos, planejamento, qualidade, entre outros. Portanto, apesar desse assunto não ser amplamente pesquisado na engenharia de software, estudos foram ou estão sendo direcionados neste assunto.

No trabalho de Menolli (MENOLLI, 2013) são apresentados vários trabalhos de engenharia de software e como estes trabalhos estão sendo aplicados na aprendizagem organizacional. Menolli (MENOLLI, 2013) faz uma classificação destes trabalhos com base no SWEBOK (IEEE, 2004). O SWEBOK divide a engenharia de software em dez áreas de conhecimento mais uma disciplina relacionada à engenharia de software.

Dentre estes estudos, nota-se uma concentração de trabalhos nas áreas de processos de engenharia de software, como por exemplo (JARRAHI; KANGAVARI, 2012) e (KIMMERLE, 2010), e gestão de engenharia de software como (CHEN, 2005) e (BOH, 2007), o que é considerado normal, pois estas áreas cobrem todo o ciclo de vida do processo de software.

Percebe-se também que o uso de ferramentas e tecnologias semânticas, a fim de aprimorar o aprendizado em projetos de engenharia de software, vem ocorrendo, o que tem sido chamado de aprendizagem organizacional semântica. Alguns estudos apresentam ferramentas que são capazes de auxiliar a troca de conhecimento e a comunicação entre os membros da organização, e assim auxiliar a propagação do conhecimento. Muitas destas ferramentas utilizam tecnologias semânticas, baseadas na proposta da Web Semântica de Berners-Lee (BERNERS-LEE; HENDLER; LASSILA, 2001). Para Berners-Lee, Hendler e Lassila (BERNERS-LEE; HENDLER; LASSILA, 2001), Web Semântica não é separada da Web, mas uma extensão dela, na qual é dado à informação um significado bem definido, permitindo que computadores e pessoas trabalhem em cooperação.

As técnicas, conceitos e aplicações da Web Semântica vêm sendo utilizadas em diversas áreas, fazendo com que novas linhas de pesquisa sejam criadas. Em Capuano, Miranda e Orcioli (CAPUANO; MIRANDA; ORCIOLI, 2009), a tecnologia da Web Semântica é aplicada em sistemas educacionais e também em sistemas de aprendizagem organizacional semânticos.

Portanto, a aprendizagem organizacional semântica pode ser considerada a aplicação de recursos da Web Semântica para promover a aprendizagem organizacional dentro de uma empresa (MENOLLI; MALUCELLI; REINEHR, 2011). Para Sicilia e Lytras (SICILIA; LYTRAS, 2005), a tecnologia da Web Semântica deve ser aplicada não só para melhorar os processos de



aprendizagem, mas como uma ferramenta de propósito para conduzir mudanças no comportamento.

Por fim, Menolli (MENOLLI, 2013) afirma que conceitos advindos da área educacional não são frequentemente utilizados para prover aprendizagem organizacional na área de engenharia de software. Dessa forma, a próxima seção introduz alguns conceitos, como objetos e unidades de aprendizagem, que são da área educacional e que são elementos utilizados no desenvolvimento do ambiente colaborativo semântico proposto em (MENOLLI; MALUCELLI, REINEHR, 2011) que serve como base para este trabalho.

## **2.3 - Ambientes e Modelos voltados à Aprendizagem Organizacional**

Muitos estudos são dirigidos à área de aprendizagem em ambientes corporativos com intuito de identificar fatores que possam auxiliar em situações específicas, gerenciando o conhecimento para facilitar a aprendizagem e o reúso do conhecimento (MENOLLI, 2013).

Na literatura são encontrados muitos trabalhos que buscam apoiar a aprendizagem organizacional principalmente em empresas desenvolvedoras de software, tendo um maior destaque a partir do final da década de 90. Primeiramente, muitas pesquisas eram feitas no sentido de criar ferramentas ou modelos baseados em conceitos de fábricas de experiências e lições aprendidas. No entanto, com o passar dos anos, outros conceitos foram integrados nesta linha de pesquisa com finalidade de melhorar a aprendizagem organizacional, principalmente com o surgimento das tecnologias *web*, como *wikis* e tecnologias semânticas.

Em Menolli (MENOLLI, 2013), são listados alguns dos trabalhos existentes na literatura, cujo objetivo é apoiar a aprendizagem na área de engenharia de software. Estes trabalhos foram divididos em vários tópicos, que são os seguintes:

- Aprendizagem colaborativa e compartilhamento de conhecimento;

- Aquisição de conhecimento integrada ao ambiente de desenvolvimento;
- Gestão do conhecimento e lições aprendidas no desenvolvimento de software;
- Modelos para aprendizagem em organizações; e
- Utilização de recursos semânticos;

Dentre estes trabalhos, alguns fazem o uso de ferramentas Web, como wiki, blogs e redes sociais, como meios de comunicação entre os membros, como (CAPUANO; MIRANDA; ORCIUOLI, 2009) e (CARRERAS et al., 2011). Capuano, Miranda e Orciuli (CAPUANO; MIRANDA; ORCIULI, 2009) e Carreiras et al. (CARREIRAS et al., 2011), também apontam estas ferramentas como meio de substituir as intranets das empresas e desta forma prover um ambiente em que a comunicação e a colaboração dos trabalhadores aconteça de maneira eficiente, oferecendo um ambiente colaborativo no qual a aprendizagem organizacional seja possível.

Outra tendência observada dentre estes trabalhos é a organização da informação de forma a auxiliar a aprendizagem por meio da utilização de recursos semânticos, como apresentado nos trabalhos de (CHOU et al., 2009) e (CAPUANO; MIRANDA; ORCIUOLI, 2009). No entanto, este tipo de abordagem é muito aplicado em ambientes educacionais, sendo muitas vezes negligenciado nos ambientes organizacionais.

## **2.4 - Representação do Conteúdo**

Nos últimos anos a aprendizagem à distância vem se tornando mais popular. Devido a este fato, a produção de materiais de aprendizagem eletrônica também aumentou (MENOLLI, 2013). Muitos destes materiais são específicos para uma única plataforma ou tecnologia, não sendo possível seu reúso. Desta forma, visando resolver este problema, têm surgido há vários anos, padrões internacionais para a produção e compartilhamento de conteúdo instrucional, que propõem principalmente a reutilização dos conteúdos.

Organizações de padronização como IEEE (*Institute of Electrical and Electronics Engineers*) e ISO (*International Organization for Standardization*), que têm como objetivo dar suporte à catalogação de conteúdo instrucional para que possam ser apropriadamente recuperados e reutilizados, criaram grupos de trabalho onde são elaboradas propostas para estruturação e categorização desses conteúdos (metadados) (BRASIL, 2007). Assim, os padrões representam os elementos para fornecer comunicação entre diferentes ambientes computacionais, bem como sua aquisição e facilidade, assim como sua interoperabilidade.

Para retratar estes elementos, alguns padrões têm sido propostos (1484.12.1 *Standard for Learning Object Metadata* e SC 36 WG 2 – *Information Technology for Learning, Education, and Training*). Entre os principais elementos encontrados na literatura, são destacados os Objetos de Aprendizagem (OAs). Em IMS (IMS, 2003), objetos de aprendizagem são definidos como qualquer recurso reprodutível e endereçável digital, ou não digital, usado para executar atividades de aprendizagem ou atividades de apoio. Outra definição é dada por Polsani (POLSANI, 2004), que afirma que um objeto de aprendizagem é uma unidade de conteúdo independente e autônoma, que pode ser reutilizada em vários contextos de ensino.

A partir de objetos de aprendizagem é possível definir outros elementos que são abordados neste trabalho, como as unidades de aprendizagem, que é um termo abstrato para se referir a qualquer parte previamente delimitada de educação ou conhecimento, como por exemplo, um curso, um módulo ou uma lição (BRASIL, 2007).

As unidades de aprendizagem representam mais do que uma coleção de artifícios ordenados para aprendizagem, uma unidade de aprendizagem inclui uma variedade de atividades prescritas (IMS, 2003).

Outro elemento que foi utilizado neste trabalho, e deve ser propriamente definido é o *Learning Design* (LD), que descreve o método que permite aos alunos atingir objetivos de aprendizagem após a realização de um conjunto de atividades, utilizando mecanismos de um ambiente (AMORIM et. Al., 2006). Para especificar os *Learning Design* são necessárias Linguagens de Modelagem Educacional (LME), que são modelos de informação semântica e

agregações que descrevem, do ponto de vista pedagógico, o conteúdo bem como as atividades educacionais (AMORIM, et al., 2006).

Para que estes elementos de aprendizagem eletrônica sejam bem definidos, é necessário ter uma forma concisa de padronizá-los e de representá-los, utilizando os conceitos apresentados. Assim, foram introduzidos padrões para definição de elementos educacionais, com a finalidade de representar uma forma de organizar os dados destes elementos para que ocorra a comunicação entre diferentes ambientes computacionais, bem como seu acesso e usabilidade, e também garantir sua interoperabilidade (DIAS et al., 2009).

De acordo com Dias et al. (DIAS et al., 2009), estes padrões são divididos de acordo com as suas funcionalidades: metadados, empacotamento, interface e comunicação, e integração.

#### **2.4.1. Unidades de Aprendizagem**

Uma unidade de aprendizagem define um módulo geral para um processo de aprendizagem, como um curso, por exemplo, (IMS, 2003). Dentre as composições da unidade de aprendizagem está o *learning design*.

##### Learning Design

A estrutura de uma unidade de aprendizagem é definida utilizando alguma Linguagem de Modelagem Educacional (LME), que são modelos de informação semântica e agregações. Estas linguagens descrevem, de um ponto de vista pedagógico, o conteúdo bem como as atividades educacionais (AMORIM et al., 2006).

As LMEs são organizadas em unidades de estudo com o objetivo de permitir a sua reutilização e interoperabilidade (RAWLINGS et al., 2002). Além disso, as LMEs facilitam a descrição de aspectos pedagógicos que estão relacionados com OAs em processos educativos (KOPER, 2001).

Uma das principais linguagens de modelagem educacional é o IMS *Learning Design* (IMS LD) (IMS, 2003), que dá suporte ao uso de diferentes

abordagens de ensino/aprendizagem, tais como: comportamentalistas, cognitivistas e construtivistas.

O IMS *Learning Design* foi originalmente criado após uma ampla avaliação e comparação entre as diversas abordagens pedagógicas e suas atividades de aprendizagem, buscando alcançar um meio termo entre aplicação pedagógica e um bom nível de generalização (BRASIL, 2007). O IMS *Learning Design* é baseado na ideia de que existem mais relações no processo de ensino e aprendizagem, do que somente a relação de um único aluno diretamente com o conteúdo (BRASIL, 2007).

Assim, a especificação IMS LD é uma metalinguagem que descreve todos os elementos do projeto de um processo de ensino e aprendizagem, e elaborado pelo grupo de trabalho IMS/GLC (IMS, 2003). IMS LD descreve um método composto por uma série de atividades conduzidas tanto pelo aluno quanto pela equipe, a fim de alcançar os objetivos de aprendizagem (AMORIM et al., 2006). Na Figura 2-2, é apresentado o modelo conceitual do IMS LD.

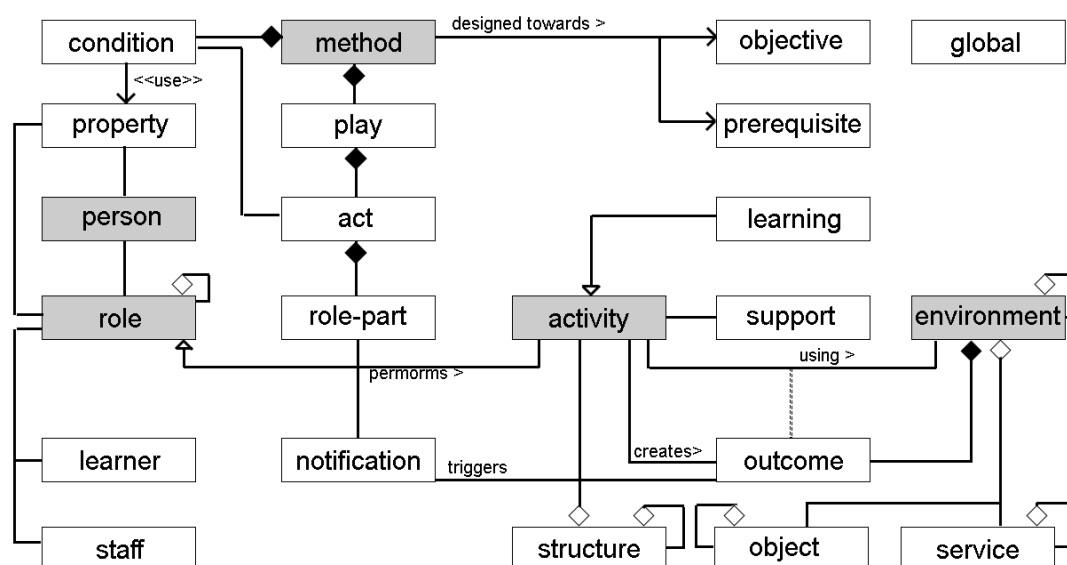


Figura 2-2. Modelo conceitual do IMS-LD (IMS, 2003)

### 2.4.2. Content Packages

O propósito do *Content Package* é prover uma forma padronizada para a troca de conteúdo de aprendizagem entre diferentes sistemas e ferramentas.

O *Content Package* também provê um local para se declarar a estrutura e o comportamento esperado de uma coleção de conteúdos de aprendizagem (ADL, 2004).

O *Content Package* desenvolvido pelo *IMS Global Learning Consortium* consiste em um padrão que realiza o armazenamento dos múltiplos artefatos digitais em um único arquivo, preservando sua organização. Assim, os *Content Packages* são a estrutura física dos cursos em uma unidade de aprendizagem, e o *IMS Content Packaging Specification* descreve como recursos digitais podem ser organizados dentro de uma unidade lógica de aprendizagem, chamada de *Content Package* (IMS, 2004). A estrutura do *IMS Content Package* (IMS CP) é mostrada na Figura 2-3.

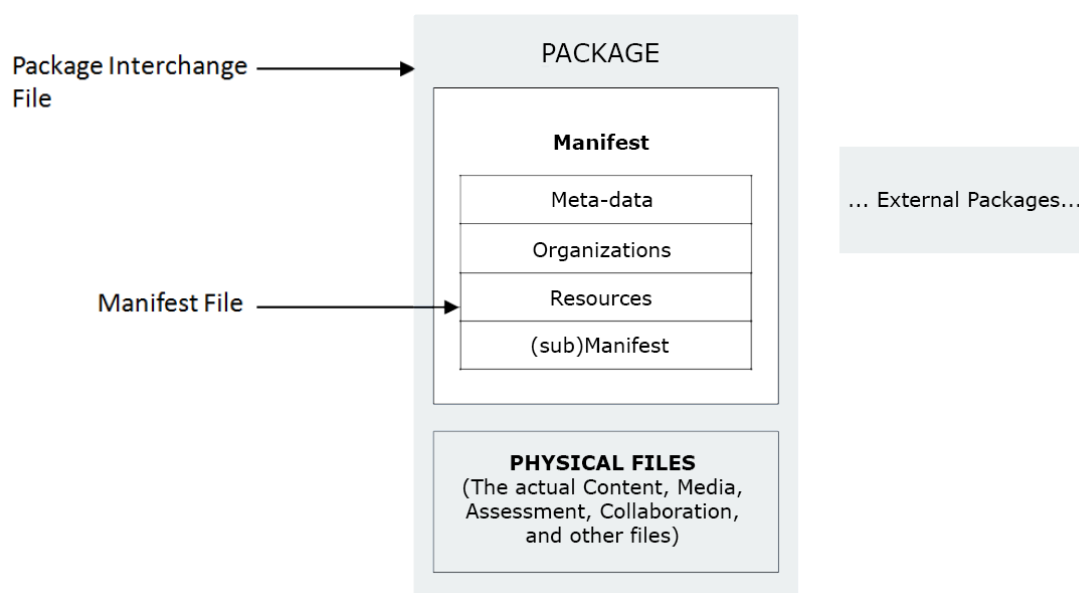


Figura 2-3. Escopo do IMS Content Packaging. (IMS, 2004)

O pacote gerado por meio da utilização do padrão IMS CP é indicado na Figura 3-3 como *External Package*. Dentro deste pacote há um arquivo denominado manifesto (*manifest*) subdividido em quatro partes (COSTA et al., 2008). O Quadro 2-1 descreve esta divisão.

Quadro 2-1. Manifesto do Padrão IMS CP. Adaptado de (COSTA et al., 2008)

Subdivisões	Descrição
Metadado	Descreve os dados do objeto.
Organização	Associa ao conteúdo um índice de tópicos e sub-tópicos formando uma

---

	estrutura hierárquica organizacional.
Recursos	Contém a relação de dependência entre os artefatos digitais que estão armazenados no <i>external packages</i> .
Sub-manifesto	Este item é opcional, caso possua pacotes dentro de pacotes.

---

No *External Package* são armazenados todos os artefatos pertencentes ao pacote. Assim, com este padrão é possível transportar para diferentes locais o pacote, que é composto por diversos arquivos.

## 2.5 - Adaptação de Padrões de Representação de Conteúdos para o Ambiente Colaborativo Semântico

No ambiente proposto por Menolli; Malucelli e Reinehr (MENOLLI; MALUCELLI; REINEHR, 2011), é proposto que diferentes ferramentas sejam integradas e organizadas utilizando uma representação comum. Esta representação permite que os diferentes tipos de ferramentas colaborativas sejam integradas e organizadas para projetos da área de engenharia de software. Para tanto, foi definida uma ontologia que guia toda a organização das ferramentas, apresentando um vocabulário comum, o qual permite que essas ferramentas sirvam como auxiliares na aquisição de conhecimento, de forma que todo o conhecimento seja organizado para as necessidades específicas de empresas de desenvolvimento de software.

Apesar dos padrões apresentados na Seção 2.4 servirem como base para a ontologia proposta por Menolli et al. (MENOLLI et al., 2014), o XML do Esquema de Aprendizagem possui algumas limitações, como (GIL; RATNAKAR, 2002):

- As relações hierárquicas entre dois ou mais conceitos não podem ser explicitamente definidas.
- As propriedades das relações não podem ser definidas. O XML do Esquema de Aprendizagem não fornece primitivas para representar nem propriedades matemáticas (como simetria ou transitividade), nem propriedades taxonômicas (como disjunções e partições exaustivas) de uma relação.

- Restrições gerais e formais (ou axiomas) entre conceitos, atributos e relações não podem ser especificadas.

Assim, uma ontologia, denominada *Unit of Organizational Learning Ontology* (UOLO) (MENOLLI et al., 2014), foi definida para orientar a organização de materiais produzidos em diferentes ferramentas colaborativas.

Nesta ontologia (UOLO) um modelo, baseado no *IMS Learning Design* (IMS, 2003), foi realizado. Este modelo se concentra principalmente na estrutura de uma Unidade de Aprendizagem. No entanto, o *IMS Learning Design* é muito mais complexo do que apenas uma forma de organizar o material na sequência correta, portanto, uma adaptação do IMS LD foi necessária para ser viável em um ambiente organizacional.

Todo conceito envolvido na dinâmica do processo de aprendizagem (*Method, Play, Act, e Activity*) estabelece uma relação com uma das subclasses do conceito *Complete Unit*, o que indica quando uma execução está terminada.

As principais diferenças entre o *learning design* proposto em Menolli et al. (MENOLLI et al., 2014), e o original IMS LD, é que na ontologia proposta os componentes relacionados com o tempo e controle da execução de uma unidade de aprendizagem não foram utilizados.

Baseado no modelo formal do *Learning Design* apresentado na Figura 2-4, que apresenta o modelo conceitual do IMS LD, um diagrama de classes da adaptação realizada no *Learning Design* é apresentado. Este modelo adaptado é apresentado na Figura 2-4.



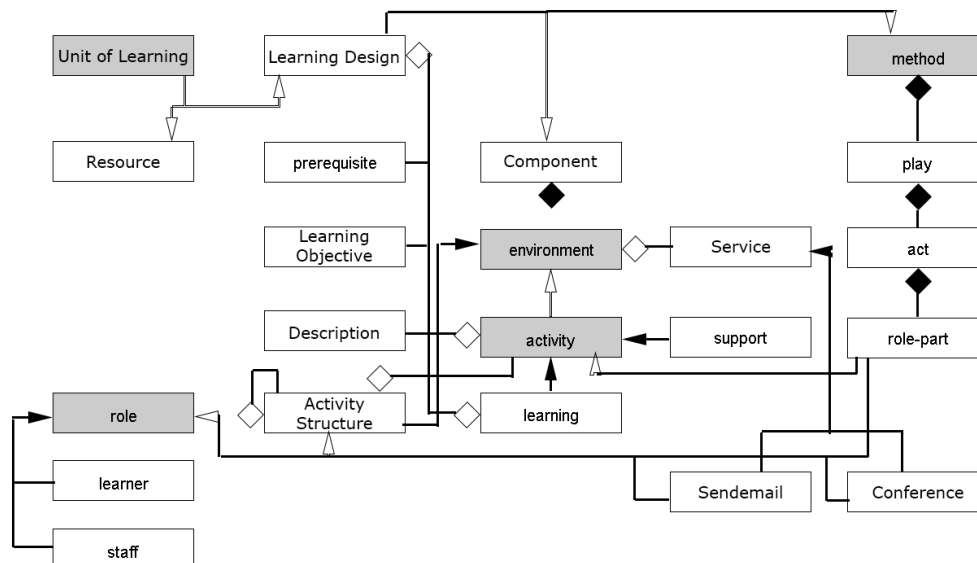


Figura 2-4. Diagrama de Classes do Learning Design da UOLO. Adaptado de (MENOLLI et al., 2014).

Também foi definido dentro da UOLO o *Content Package*. O *Content Package* descreve a estrutura física de um curso definido pelo *learning design*. Menolli (MENOLLI, 2013), utilizou o IMS Content Packaging Specification (IMS, 2004) para definir os conceitos envolvidos no incremento de Content Packages.

Após a definição do conjunto de propriedades a serem usadas na ontologia, um modelo formal foi definido. O modelo é apresentado na Figura 2-5.

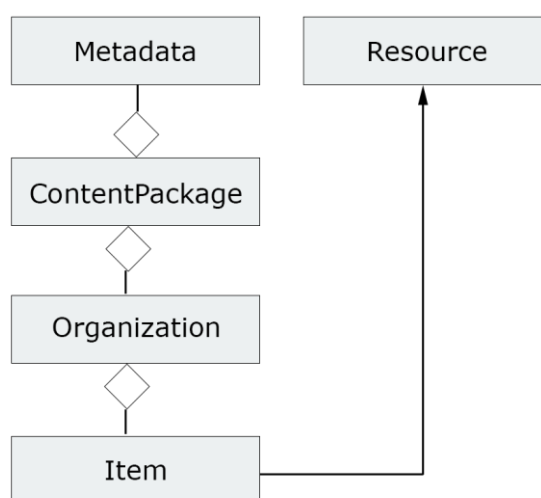


Figura 2-5. Diagrama de Classes do Content Package da UOLO.

## 2.6 - O Ambiente Colaborativo Semântico

O propósito do ambiente é que diversos conteúdos, tais como textos, imagens, mídias, artefatos ou código fontes, sejam inseridos em alguma ferramenta social, como por exemplo, *wiki*. A partir dos conteúdos destas ferramentas, são gerados então, objetos de aprendizagem organizacionais e posteriormente unidades de aprendizagem organizacionais que são compostas por um ou mais *Content Packages*. A Figura 2-6 apresenta o processo de criação das unidades de aprendizagem organizacionais.

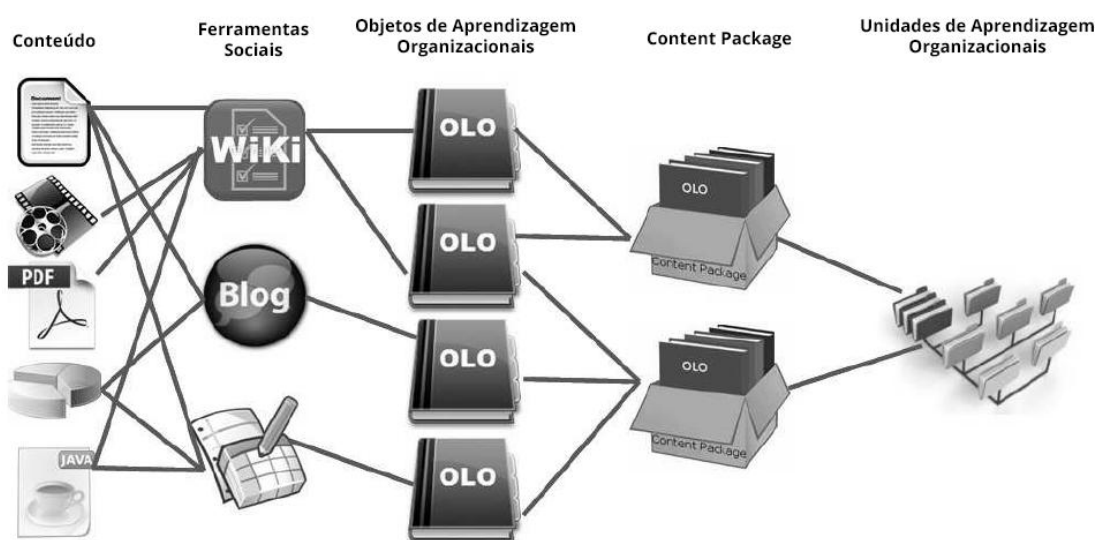


Figura 2-6. Esquema para Geração de Unidades de Aprendizagem Organizacionais (MENOLLI, 2013).

A implementação do ambiente se deu a partir da arquitetura proposta em (MENOLLI; MALUCELLI; REINEHR, 2011), porém, nesta implementação, apenas a tecnologia *wiki* foi utilizada na camada de aplicação. Menolli (MENOLLI, 2013) atribui esta escolha ao fato da *wiki* contribuir para socialização, e sendo integrada ao ambiente, auxilia a promover a externalização e internalização do conhecimento (KIMMERLE; MOSKALIUK. CRESS, 2009), bem como ao consenso existente entre as empresas desenvolvedoras de software no sentido de ter a *wiki* como a melhor tecnologia para o compartilhamento do conhecimento, além de ser a tecnologia que melhor pode auxiliar a promover conceitos das teorias estudadas de

aprendizagem e compartilhamento de conhecimento (MENOLLI; MALUCELLI; REINEHR, 2014).

O ambiente permite que a partir de conteúdos inseridos pelos usuários na ferramenta social, sejam criados objetos e unidades de aprendizagem organizacionais semiautomaticamente, como é mostrado na Figura 2-7. O esquema é dividido em três partes principais, que são pré-processamento, gerador do curso e busca.

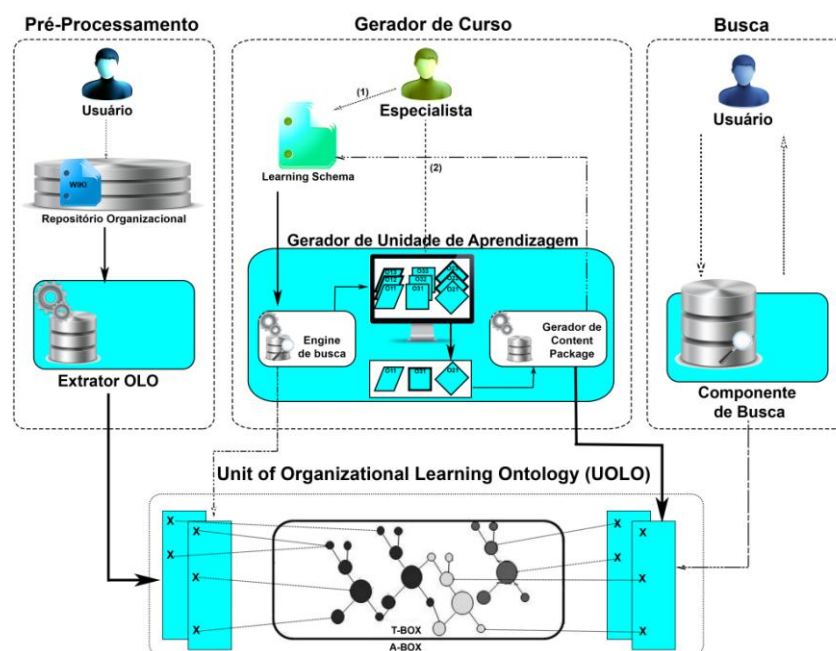


Figura 2-7. Esquema de Ambiente Colaborativo Semântico (MENOLLI, 2013).

### 2.6.1. Pré-processamento

O pré-processamento é a primeira etapa para gerar as unidades de aprendizagem organizacionais. A inserção dos conteúdos na *wiki* acontece nesta etapa e a partir destes conteúdos inseridos que serão gerados os Objetos de Aprendizagem Organizacional (OLO – *Organizational Learning Object*).

As *wikis* são definidas a partir de informações instrucionais e de contexto da área da engenharia de software contidas em um metamodelo chamado *Meta model of Instructional Context (MIC)*, que tem como objetivo organizar as

ferramentas para promover a integração da informação, respeitando o contexto do domínio do conhecimento, bem como o papel instrucional dos conteúdos.

Por meio de uma *wiki* privada denominada *Semantic Software Engineering Wiki* (SSEW), os documentos *wiki* serão configurados conforme a MIC.

### **2.6.2. Gerador de Curso**

O gerador do curso é a etapa responsável por gerar os *Content Packages* que compõe as unidades de aprendizagem organizacionais. Em um primeiro momento, um especialista é responsável por criar um Esquema de Aprendizagem, que define a estrutura da unidade de aprendizagem.

Após a definição do *Learning Schema*, as unidades de aprendizagem são geradas por meio do Gerador de Unidade de Aprendizagem.

#### Gerador de Unidade de Aprendizagem

Menolli (MENOLLI, 2013) propõe um Gerador de Unidade de Aprendizagem no qual é dividido em dois subcomponentes, a *Engine* de Busca e o Gerador de *Content Package*.

A *Engine* de Busca lê a estrutura do curso definida no Esquema de Aprendizagem e procura na UOLO por OLOs que sejam candidatos para cada uma das atividades de aprendizagem definidas para o curso. Assim, a *Engine* de Busca retorna uma lista de OLOs candidatos para cada *learning activity*. Posteriormente, caberá ao especialista escolher nesta lista, o melhor objeto de aprendizagem organizacional para cada atividade de aprendizagem.

Uma vez selecionados os OLOs para cada *learning activity* definida no *Esquema de Aprendizagem*, o Gerador de *Content Packate* acessa o *Esquema* para criar os *Content Packages*.

### **2.6.3. Componente de Busca**

Após os objetos e unidades de aprendizagem serem inseridos na base de conhecimento da organização, os usuários poderão realizar pesquisas por

objetos de aprendizagem organizacionais ou unidades de aprendizagem organizacionais.

Embora na arquitetura apresentada na Figura 2-7 esteja previsto um componente de busca semântica, o autor optou pelo desenvolvimento do primeiro protótipo do ambiente, a busca sem suporte semântico. Este componente de busca procura na ontologia, tanto por objetos, como por unidades de aprendizagem organizacionais, de acordo com uma palavra-chave que é definida pelo usuário no momento da busca.

## **2.7 - Educação em Organizações e Design Instrucional**

A área de educação sofreu ao longo dos anos, diversas mudanças de paradigmas. Reigeluth (REIGELUTH, 1999), aponta que o paradigma de aprendizagem evoluiu de métodos cognitivistas, os quais são orientados aos objetivos de aprendizagem, para métodos voltados ao desenvolvimento de competências, considerados os modelos construtivistas.

Isto ocorreu devido a mudanças fundamentais dentro das organizações, como a retirada de processos burocráticos, controles centralizados, relações adversárias para substituição por organizações baseadas em equipes, redes de comunicação e decisões compartilhadas (REIGELUTH, 1999).

Estas mudanças em sistemas empresariais tiveram implicações importantes para a criação de instruções. Colaboradores precisam estar aptos a pensar e resolver problemas, trabalhar em equipe, exercer a comunicação, tomar iniciativas, e trazer diversas perspectivas para o seu trabalho, demonstrando o impacto do aprendizado para os objetivos estratégicos da empresa (HEQUET, 1995).

Para atingir este objetivo, é necessário que os colaboradores possuam meios para aprender e se desenvolver. Em sua obra, Reigeluth (REIGELUTH, 1999) descreve as teorias e modelos de *design* instrucional, que oferecem guias explícitos em como melhor auxiliar pessoas a aprender.

Estas abordagens teóricas denominadas de Teorias e Modelos de Design Instrucional são definidas por Reigeluth (REIGELUTH, 1999) como os

processos complexos ao nível de comunicações humanas nos quais há o inter-relacionamento de conteúdos com objetivo de auxiliar pessoas a aprender e a se desenvolver.

A ideia básica de inter-relacionamento de conteúdo é a forma como os elementos estão estruturados (BRUNER, 1966). Wilson e Cole (WILSON; COLE, 1992) assumiram a noção de conteúdo estruturado como um modelo cognitivo estruturado. A partir de modelos como este, também propostos por Reigeluth et al., (REIGELUTH, 1979), o entendimento de estruturação permitiu a criação de abordagens teóricas capazes de interrelacionar conteúdos para elaboração de lições, cursos e unidades.

A partir destas linhas de estudo, surgiram algumas abordagens com objetivo de auxiliar na estruturação do conhecimento, como a Teoria da Elaboração (*Elaboration Theory*), que é um modelo para sequenciar e organizar cursos de uma instrução (REIGELUTH et al., 1979). De acordo com a teoria da elaboração, uma instrução deve ser organizada de forma crescente de complexidade para melhor aprendizado. A teoria de elaboração propõe sete componentes estratégicos para elaboração de cursos:

1. Uma elaboração sequencial;
2. Sequencias de pré-requisitos de aprendizagem;
3. Sumário;
4. Síntese;
5. Analogias;
6. Estratégias cognitivas; e
7. Controle do aprendiz.

O primeiro componente é o mais importante de acordo com a teoria da elaboração. É dito que a abordagem proposta pela teoria da elaboração, resulta em uma formação de estruturas cognitivas estáveis e, portanto, ocasiona melhor retenção e transferência do conhecimento, melhoria da motivação do aprendiz por meio da criação de contextos significativos e o fornecimento sobre o conteúdo que permite melhor controle do aprendiz.

O que se pode observar desta teoria da elaboração apresentada por Reigeluth e Stein (REIGELUTH; STEIN, 1983), é que este é um modelo cognitivista orientado a objetivos de aprendizagem. O modo como a educação

é tratada, mudou com passar dos anos, principalmente dentro de organizações o modelo de aprendizagem é voltado ao desenvolvimento de competências, sendo considerado um modelo construtivista (REIGELUTH, 1999).

Baseado neste modelo, Reigeluth (REIGELUTH, 1999) realiza uma nova elaboração de sua teoria, no entanto, com foco no ponto inicial, a Sequência da Elaboração. A teoria da elaboração faz uma distinção entre competências em execução de tarefas e competências em assuntos ou domínios. Competências em execução de tarefas são relacionadas ao aprendiz se tornar um especialista em uma habilidade específica, como gerenciar um projeto, vender um produto ou escrever um plano anual. Competências em domínios estão relacionadas a assuntos não ligados a uma habilidade, como economia, programação, eletrônica, etc, nos quais o aprendiz se tornará um especialista.

No desenvolvimento de competências em habilidades, uma tarefa pode variar de ações simples para ações complexas. O foco da teoria de elaboração é auxiliar no desenvolvimento do aprendiz em execução de tarefas mais complexas. Para auxiliar na aprendizagem de tarefas complexas, a teoria da elaboração oferece o SCM (Simplifying Conditions Method) para delinear uma sequência de apresentação “simples para o complexo”, iniciando com conteúdos mais simplistas e exemplos da tarefa, e gradualmente progredindo para conteúdos mais complexos.

O método SCM oferece um guia prático para analisar, selecionar e sequenciar conteúdo. Para definir uma sequência de conteúdos com método SCM é necessário primeiramente que a tarefa que se deseja construir seja analisada. Esta análise é baseada em perguntas que são realizadas com o objetivo de entender como a habilidade será construída, por exemplo: “Qual é a forma mais simples em que esta tarefa foi realizada?”; e “Qual é a próxima forma mais simples”, e assim por diante. A cada versão identificada, os conteúdos são determinados.

Portanto, o método SCM define três diferentes fases de definição de conteúdos:

- Fase 1: Preparação para Análise e Definição. Nesta fase são identificadas as características gerais da tarefa que se deseja especializar um colaborador.

- Fase 2: Identificação do episódio mais simples da tarefa. Nesta fase, a forma mais simples de como a tarefa é executada é definido.
- Fase 3: Identificação do Próximo episódio. Nesta fase, a próxima forma mais simples da tarefa é identificada.

Por fim, estas versões são organizadas a partir dos assuntos mais simples para os mais complexos e apresentados ao aprendiz. No entanto, o aprendizado construtivista depende da ativação de vários processos cognitivos no aprendiz durante o seu processo. Além da seleção de informação relevante apresentada pela Teoria da Elaboração (REIGELUTH, 1999), também dependendo da organização da informação selecionada, e integração da informação organizada com o conhecimento existente do aprendiz. Mayer (MAYER, 1996) define esta análise ao modelo SOI, que identifica três processos cruciais para a aprendizagem construtivista: S de seleção; O de organização; e I de integração.

No modelo SOI, o primeiro processo, de seleção, tem como objetivo a seleção de informação relevante para o aprendizado. Segundo Reigeluth (1999) quando informações são apresentadas ao aprendiz, a forma resumida desta informação é facilmente relacionada. Dada a capacidade limitada do ser humano processar informações, apenas algumas destas informações podem ser retidas para futura aplicação no trabalho. Por este motivo, é importante que apenas informações relevantes sejam apresentadas no processo de aprendizagem.

O segundo processo, de organização, envolve estruturar as representações selecionadas em uma representação coerente. Kintsh (KINTSH, 1988) referencia esta atividade como construir uma situação modelo a partir da informação apresentada.

O terceiro processo do modelo SOI, de integração, é onde os aprendizes irão criar conexões entre os elementos selecionados e organizados com o seu conhecimento atual.



## **2.8 - Considerações Finais**

O presente capítulo apresentou os principais conceitos aplicados ao longo deste estudo como a aprendizagem organizacional, padrões para representação de conteúdos que são elementos fundamentais para entender o funcionamento do ambiente colaborativo semântico. Também foi apresentado neste capítulo o estado atual dos trabalhos relacionados à área de aprendizagem organizacional, com foco na aprendizagem organizacional na área da engenharia de software. O capítulo também apresentou o ambiente colaborativo semântico para empresas desenvolvedoras de software, implementado em (MENOLLI, 2013), bem como sua arquitetura e os elementos que a compõe, que são o pré-processamento, o gerador de curso e o componente de busca. Por fim, foram apresentados alguns conceitos sobre a área pedagógica, mais especificamente sobre modelos de design instrucional, bem como um paradigma atual sobre a educação em organizações.

## Capítulo 3 - **Descoberta de Conhecimento Baseadas em Texto**

Neste capítulo é apresentada uma visão geral teórica sobre os conceitos que são abordados neste trabalho, como motores de busca, mineração de textos (*text mining*) bem como conceitos do domínio de sistemas de recomendação (*Recommendation Systems*), sendo discutidas algumas definições e técnicas que são utilizadas para recuperar e tratar informações, assim com alguns tópicos relacionados ao seu desenvolvimento.

### **3.1 - Introdução**

A velocidade e a amplitude com que o conhecimento gerado passou a ser compartilhado provocaram o surgimento de uma dinâmica de reaproveitamento e produção de novos conhecimentos, bem como o aparecimento de novas necessidades de tratar a informação. Para suprir essas necessidades, ferramentas e técnicas para tratamento de grandes massas de dados foram criadas e aperfeiçoadas, como por exemplo, o tratamento estatístico aplicado no processamento e análise de dados e informações. As técnicas de tratamento de dados busca de retratar o que não seria possível com a limitada capacidade humana de leitura e registro (RAMOS, BRÄSCHER, 2009).

No entanto, após o surgimento das ferramentas Web 2.0 (Motores de Busca, *Wikis*, *Folksonomias*, entre outros), a Web (*World Wide Web*) se tornou o meio popular e interativo de disseminar a informação. A Web é considerada diversa, dinâmica e sem limites. Isso contribuiu para o surgimento da escalabilidade, dados em multimídia, e questões relacionadas ao tempo de espera para conseguir informação (KOSALA; BLOCKEEL, 2000). Algumas destas questões levantadas levaram a evolução de tópicos como a descoberta de informação relevante, criação de novos conhecimentos a partir da informação disponível na internet, personalização da informação e aprendizado sobre usuários individuais.

A descoberta de conhecimento nos dados originados da internet (também conhecido como *Web Mining*) tem sido uma área de pesquisas interdisciplinares de recente interesse. Para a descoberta de padrões, o uso de *Web Mining*, que é a mineração de arquivos de log de servidor entre outras técnicas, provou ser o paradigma mais apropriado para descoberta de conhecimento na internet.

Nas últimas décadas, várias abordagens foram propostas com o objetivo de contribuir com a área de descoberta de conhecimento na internet. Algumas delas apresentam ferramentas para processamento de tarefas de *web mining* como descoberta de regras de associação e padrões sequenciais de dados da internet (COOLEY et al., 1999). Outras apresentam técnicas de *web mining* para arquivos de log de servidor e consolidação dos dados (ZAIANE et al., 1998).

Várias técnicas de diferentes áreas de pesquisa, como Bancos de Dados, Recuperação de Informação e Processamento de Linguagem Natural também têm sido utilizadas. Isso ocorre, principalmente, porque nos últimos anos, os focos da pesquisa têm se voltado para aqueles conteúdos armazenados em meio digital, sem a preocupação com o rigor da estruturação – os documentos textuais – comumente chamados de “informação não estruturada”. Esses conteúdos se revelaram portadores de informações valiosas, camufladas em grandes volumes textuais, que passaram a ser explorados em busca de padrões de conhecimento até então desconhecidos, para tomada de decisão e geração de novos conhecimentos (RAMOS, BRÄSCHER, 2009).

Enquanto que aplicações clássicas têm focado em processamento e mineração de texto, o advento da internet possibilitou a necessidade de novos métodos para mineração e aplicação (AGGARWAL; ZHAI, 2012). As próximas seções deste capítulo apresentam detalhadamente as técnicas e meios em que as técnicas de *Web Mining* têm sido utilizadas e que de alguma forma contribuíram para o desenvolvimento deste trabalho.

### 3.2 - Motores de Busca

O foco do uso de motores de busca neste trabalho é sua utilização para recuperar os interesses de desenvolvedores de software em uma organização. Desta forma, não se faz necessário apresentar como é o funcionamento de um motor de busca e as diversas atividades que são executadas de modo a retornar a pesquisa realizada, uma vez que estes conceitos não foram explorados no desenvolvimento do trabalho. No entanto, é necessário explorar como o uso de motores de busca pode auxiliar na coleta de interesse de desenvolvedores de software.

Nos últimos anos, muitos trabalhos foram realizados com objetivo de auxiliar desenvolvedores de software na busca por códigos fonte. Isto porque desenvolvedores estão constantemente buscando por código fonte que possam ser reutilizados em projetos, ou como um exemplo ou referência (VALENCIA; SIM, 2009).

Este uso constante de motores de busca por desenvolvedores de software produz circunstâncias nas quais é possível atuar de forma a colaborar com a aprendizagem organizacional voltada para empresas desenvolvedoras de software. Não é necessário criar um ambiente específico, que possua um motor de busca integrado para poder coletar as consultas de um programador. É possível aproveitar o uso constante destas ferramentas em prol da descoberta de conhecimento.

A década de 90 foi estabelecida como um marco para a Web (*World Wide Web*), radicalmente mudando a disponibilidade de informação e tornando possível a rápida disseminação de informação digital através do mundo (LEVENE, 2011). Os motores de busca agora indexam bilhões de páginas e este número é considerado apenas uma fração do total de informação que pode ser acessada na Web (LEVENE, 2011). A maioria dessa informação pode ser acessada não diretamente, mas por meio dos motores de busca. Portanto, os motores de busca são um dos caminhos primários para que usuários da Internet encontrem web sites.

No entanto, a partir dos anos 90 nota-se um desenvolvimento da internet como fonte de conhecimento. Motores de busca que foram populares

como Netscape e AOL foram substituídos por motores de busca de melhor desempenho e que fornecem múltiplas aplicações online (SEMAG, 2015).

Estudantes e profissionais de todos os níveis se voltam para a internet para pesquisar e armazenar informação. Assim, ao pesquisar sobre tópicos de interesse, os motores de busca fornecem o que parece ser uma fonte ilimitada de informação. Fato que torna relativamente simples as tarefas de recuperar definições, diagramas, artigos, estudos, opiniões, novos arquivos e pessoas com o mesmo interesse.

O termo Motor de Busca é facilmente relacionado a sites de busca online como Google, Bing e Yahoo. No entanto, existem na atualidade outros motores de busca que apresentam propostas diferentes de indexação e retorno de consultas. O Quadro 3-1 apresenta, juntamente com os motores de busca mais populares, alguns outros motores de busca que são utilizados no mundo. Alguns destes motores utilizam conceitos de Web Crawling e Indexação. O conceito de *Web Crawling* é definido pelo Google (GOOGLE, 2015) como a navegação pelos links de páginas de forma a conhecer os assuntos, enquanto que a indexação é a manutenção destas informações em um computador central. Outro conceito é o de *Meta Crawling*, que é a mescla de resultados de motores de busca populares como Google, Bing, entre outros.

A característica em comum identificada entre a maioria destes motores de busca é o chamado Método de Requisição. Na internet, quando se navega em uma página, as transferências de informações entre o servidor da página e o computador do usuário são realizadas por meio do protocolo denominado HTTP (*Hypertext Transfer Protocol*). Internamente, o protocolo pode realizar esta transferência de diferentes formas. Estas formas são chamadas de métodos de requisição, e são as seguintes: *GET*; *POST*; *HEAD*; *PUT*; *DELETE*; *OPTIONS*; *TRACE* e *CONNECT* (W3, 2015).

Dentre estes métodos de requisição, o mais utilizado pelos motores de busca é o método *GET*. O método *GET* tem como objetivo recuperar qualquer que seja a informação (páginas, figuras, documentos, etc.) utilizando uma requisição por meio de um URL (*Uniform Resource Locator*) de requisição.

Quadro 3-1. Motores de Busca mais utilizados no mundo.

Motor de Busca	URL	Indexação
<b>Google</b>	HTTP://google.com/?nord=1	Web Crawling e indexação.
<b>Bing</b>	HTTP://www.bing.com	Web Crawling e indexação.
<b>SenseBot</b>	HTTP://www.sensebot.net	Motor de busca semântico que gera um resumo de texto de várias páginas da Web sobre o tema de pesquisa. Usa mineração de texto para extrair sentido a partir de páginas da Web e apresentá-lo para o usuário de uma forma coerente.
<b>Exalead</b>	HTTP://www.exalead.com/search	Web Crawling e Indexação.
<b>Factbites</b>	HTTP://www.factbites.com	Utilizam frases digitadas pelos usuários para busca no Wikipedia.
<b>Ask</b>	HTTP://www.ask.com	Web Crawling e Indexação.
<b>AOL Search</b>	HTTP://search.aol.com/aol/webhome	Web Crawling e Indexação.
<b>Wow</b>	HTTP://www.wow.com	Web Crawling e indexação.
<b>WebCrawler</b>	HTTP://www.webcrawler.com	Web Crawling e Indexação.
<b>MyWebSearch</b>	HTTP://home.mywebsearch.com	Web Crawling e Indexação.
<b>Infospace</b>	HTTP://infospace.com	MetaCrawler.
<b>Info</b>	HTTP://www.info.com	MetaCrawler.
<b>DuckDuckGo</b>	HTTPS://duckduckgo.com	Motor de busca utiliza Crowdsourcing (Wikipedia) para melhorar a relevância dos resultados. Não registra as informações do usuário.
<b>Contenko</b>	HTTP://www.contenko.com	MetaCrawler.
<b>Dogpile</b>	HTTP://www.dogpile.com	MetaCrawler.
<b>Alhea</b>	HTTP://www.alhea.com	MetaCrawler.

### Estrutura de um URL

Um URL de requisição identifica os recursos que serão aplicados na requisição a um servidor, ou seja, a consulta informada no motor de busca é organizada de forma estruturada de modo que seja possível realizar a requisição desta consulta a um servidor. Em redes TCP/IP, um URL completo possui a seguinte estrutura:

esquema://domínio:porta/caminho/recurso?query\_string#fragmento

- Esquema é o protocolo, podendo ser HTTP, HTTPS, FTP (*File Transfer Protocol*), etc.
- Domínio é o endereço do servidor disponibilizador.
- Porta é o ponto lógico onde será realizada a conexão.
- Caminho especifica o local em um sistema de arquivos no qual se encontra o recurso, dentro do servidor.
- Query String é um conjunto de um ou mais pares de “pergunta-resposta” ou “parâmetro-argumento” (como por exemplo, nome=fulano, em que nome pode ser, por exemplo, uma variável, e fulano é o valor (argumento) atribuído a nome) (W3 SCHOOLS, 2015a). É uma string enviada ao servidor para que seja possível filtrar ou até mesmo criar o recurso.
- Fragmento é uma parte ou posição específica dentro do recurso.

Um exemplo de um URL completo de uma pesquisa realizada em um motor de busca seria:

[HTTP://www.google.com.br/?query=engenharia+de+software+pucpr](http://www.google.com.br/?query=engenharia+de+software+pucpr)

Neste exemplo o protocolo é o HTTP, o servidor é designado por [www.google.com.br](http://www.google.com.br) e a *query string* apresenta o parâmetro “query”. Cada motor de busca pode especificar seus parâmetros de forma diferente. Para este caso o parâmetro “query” tem atribuído o valor “engenharia+de+software+pucpr” que é uma pesquisa digitada no campo de busca. Ou seja, o servidor do motor de busca Google está sendo requisitado por meio de uma conexão HTTP, a localizar páginas que sejam relacionadas com a busca “Engenharia de Software PUCPR”.

Em um ambiente empresarial, com uma rede intermediada por um servidor Proxy, ao ser requisitado um URL pelo usuário, por meio do motor de busca, esta URL será registrada em um arquivo de acessos, sendo possível posteriormente saber o que foi pesquisado no motor de busca.

### 3.3 - Mineração de Texto

Na literatura pode ser encontrado diferentes definições do termo descoberta de conhecimento em banco de dados ou *knowledge discovery in databases* (KDD) e mineração de texto. De forma a distinguir mineração de texto de KDD, de acordo com Fayyad, Piatetsky-Shapiro e Smith (FAYYAD; PIATETSKY-SHAPIRO; SMITH, 1996), KDD é um processo de identificação de padrões válidos, novos, potencialmente úteis e entendíveis nos dados estruturados. Mineração de texto ou descoberta de conhecimento a partir de textos (*knowledge discovery from Text* - KDT) foi inicialmente mencionado por Feldmann, Munch e Vorwieger (FELDMANN; MUNCH; VORWIEGER, 1998).

O KDT, que lida com a análise apoiada em máquina de textos, utiliza técnicas de recuperação de informação, extração de informação bem como Processamento de Linguagem Natural (PLN) e as conecta com algoritmos e métodos de KDD, mineração de dados, entre outros (HOTHO; NUNRBERGER; PAAB, 2005).

Várias definições a respeito de mineração de texto foram apresentadas nas últimas décadas. Mineração de texto, como definida por Gao, Chang e Han (GAO; CHANG; HAN, 2005), é a descoberta de conhecimento previamente desconhecido, pela extração automática de informação de diferentes fontes. Para Fayyad, Piatetsky-Shapiro e Smyth (FAYYAD; PIATETSKY-SHAPIRO; SMITH, 1996), mineração de texto, também conhecido como mineração de dados textuais (HEARST, 1997), geralmente refere-se ao processo de extração de padrões ou conhecimento, a partir de documentos textuais.

De forma a motivar mineração de texto por perspectivas específicas, é possível atribuir para diferentes áreas, diferentes definições:

- O campo de Extração de Informação assume que mineração de texto, essencialmente corresponde a extração de fatos a partir de textos (HOTHO; NUNRBERGER; PAAB, 2005).
- Em Mineração de Dados Textuais, a mineração de texto pode ser vista como a aplicação de algoritmos e métodos de campos como



aprendizagem de máquina e estatísticas em textos, com o objetivo de encontrar informações e padrões úteis. Para isto é necessário previamente processar o texto de acordo. Muitos autores utilizam métodos de extração de informação, processamento de linguagem natural ou alguns passos de processamentos simples, de forma a extrair os dados dos textos (NAHM; MOONEY, 2002) (GAIZAUSKAS, 2003).

- Para o campo de Descoberta de Conhecimento em Bancos de Dados, seguindo o modelo de processo de descoberta de conhecimento (WIRTH, 2000), é comumente encontrado na literatura a mineração de texto como um processo com uma série de passos parciais. Hearst (HEARST, 1999) resume de forma geral como a extração de informação não descoberta em grandes bases de dados. Também Kodratoff (KODRATOFF, 1999) e Gomez (GOMEZ, 2002) consideram mineração de texto como um processo orientado a abordagens em textos.

Portanto, a mineração de texto é vista como uma extensão de mineração de dados ou descoberta de conhecimento a partir de bancos de dados estruturados. Neste trabalho, a mineração de texto é considerada principalmente como a mineração de dados textuais. Assim, o foco é em métodos que extraem padrões a partir de texto, de forma a categorizar ou estruturar coleções de texto ou extrair informação útil.

### **3.3.1. Descoberta de conhecimento a partir de texto**

Para realizar a mineração de grandes coleções de textos é necessário pré-processar os textos e armazenar a informação em uma base estruturada. Na forma estruturada, o texto descrito se torna mais apropriado para o processamento. A maioria dos métodos existentes exploram a estrutura sintática e semântica de textos. Grande parte destes métodos é baseada na ideia de que um texto pode ser representado pelo grupo de palavras que o compõe (*bag-of-words*) (HOTHO; NURNBERGER; PAAB, 2005). Por exemplo,

uma frase ou um texto composto por uma ou mais palavras, em um processo de mineração de texto, será considerado um grupo de palavras.

De forma a obter todas as palavras que são utilizadas em um texto, um processo denominado *tokenização* é necessário, no qual o texto é separado em palavras, removendo pontuações, espaços e outros caracteres não textuais. Esta representação *tokenizada* é então utilizada para processamentos futuros. O grupo de diferentes palavras obtidas pela separação dos textos, é chamado de dicionário de uma coleção.

De modo a reduzir o tamanho do dicionário, e assim a dimensão dos grupos que serão armazenados e processados, o grupo de palavras originados da *tokenização* dos textos pode ser reduzido por filtragens e métodos de *stemming* (derivação de palavras).

Os métodos de filtragem removem palavras do dicionário. Um método padrão de filtragem é a retirada de palavras que contêm pouca ou nenhuma informação (*stop words removal*), como artigos (a, o, e), conjunções (mas, então), preposições (por, de, das), etc. De forma a reduzir ainda mais o número de palavras no dicionário, métodos de indexação de termos podem ser utilizados (DEERWESTER et al., 1990) (WITTEN; MOFFAT; BELL, 1999). Neste caso, apenas palavras chaves selecionadas são utilizadas.

Os métodos de *stemming* tentam construir a forma básica de palavras, por exemplo, retirando o “s” de pronomes no plural, retirando conjugações de verbos ou outros afixos. Um grupo de derivadas são palavras com significado similar. Depois de realizar o processo de *stemming*, cada palavra é representada pela sua derivada. Alguns trabalhos apresentam algoritmos para *stemming*, no entanto, cada algoritmo compreende apenas um idioma, uma vez que cada idioma tem suas regras de escrita. Porter (PORTER, 1980) apresenta uma regra baseada em algoritmos de *stemming*, para iterativamente transformar palavras em inglês para sua forma derivada. Para o português, vários algoritmos de *stemming* foram apresentados na literatura, como por exemplo, PTStemmer (PTSTEMMER, 2015) que acompanha o código fonte para aplicação, o RSLP (RSLP, 2007) e o (ALVARES; GARCIA; FERRAZ, 2005) no qual o tratamento do sufixo é baseado e um estudo estatístico de frequência de palavras encontradas em sites brasileiros.

No entanto, a tokenização, filtragem e stemming são apenas algumas etapas do processo de mineração de texto. Normalmente, precisa-se estruturar as palavras encontradas para poder facilitar sua análise. Para isto, existem vários métodos, dentre os quais destaca-se a clusterização.

### 3.3.2. O algoritmo de clusterização CLIQUE

A principal razão para aplicação de métodos de mineração de dados em coleções de texto é com propósito de estruturá-los. Uma estruturação pode simplificar significativamente o acesso à informação importante pelo interessado. Os métodos existentes para estruturação de coleções de textos tentam automaticamente estruturá-los encontrando grupos de coleções similares (método de clusterização).

Os métodos de agrupamento (*clusterização*) podem ser utilizados de forma a encontrar grupos de textos com conteúdo similar. Cada cluster consiste em um número de documentos textuais que foram considerados similares pelo algoritmo e dissimilares de documentos de outros *clusters*. A qualidade dos *clusters* é considerada melhor se o conteúdo do documento textual, dentro de um *cluster*, é mais similar e entre *clusters* diferentes mais dissimilar.

Os algoritmos de agrupamento geram os *clusters* baseados em atributos dos dados e medidas de similaridade ou dissimilaridade. No entanto, a necessidade de diferentes aplicações de clusterização, fez com que diferentes tipos de algoritmos surgissem. Steinbach, Ertöz e Kumar (ERTOZ; KUMAR, 2003) apresentam em uma pesquisa diferentes tipos de algoritmos de agrupamento e os tipos resultantes de clusters.

Algoritmos de clusterização são divididos em uma grande variedade de diferentes tipos como aglomerativos, particionais e os algoritmos baseados em modelagem paramétrica de padrões (AGGARWAL; ZHAI, 2012). Além disso, representações textuais podem também ser tratadas como *strings* (ao invés de grupos de palavras). Estas diferentes representações necessitam de diferentes classes de algoritmos de clusterização.

Aggarwal e Zhai (AGGARWAL; ZHAI, 2012) descrevem uma lista de algoritmos de clusterização baseados em distâncias, os quais são designados

por funções de similaridade para medir a proximidade entre os objetos. Estes algoritmos são normalmente utilizados para clusterização de textos. Juntamente com alguns outros métodos coletados na literatura, esta lista é apresentada no Quadro 3-2.

**Quadro 3-2. Métodos mais utilizados para clusterização de textos (AGGARWAL; ZHAI, 2012).**

<b>Algoritmo</b>	<b>Tipo</b>	<b>Descrição</b>
<b>Ligação única</b>	Aglomerativo	Similaridade entre qualquer par de documentos de dois grupos (AGGARWAL; ZHAI, 2012).
<b>Média do grupo</b>	Aglomerativo	A similaridade entre dois clusters é a média de similaridade entre os pares de documentos entre destes dois grupos (AGGARWAL; ZHAI, 2012).
<b>Ligação completa</b>	Aglomerativo	A similaridade entre dois clusters é o pior caso de similaridade entre qualquer par de documentos entre estes dois grupos (AGGARWAL; ZHAI, 2012).
<b>Modelagem distribucional</b>	Hierárquico	Modelar a frequência da presença de palavras em documentos com o uso de distribuição multi-poisson (SAHOO et al., 2006).
<b>K-medoid</b>	Particional	Determinar um grupo otimizado de documentos representativos a partir da base original nos quais os clusters foram construídos (AGGARWAL; ZHAI, 2012).
<b>K-means</b>	Particional	Iniciar a clusterização com os grupos da base original, e então atribuir documentos a estes grupos baseados na similaridade mais próxima (AGGARWAL; ZHAI, 2012).
<b>Star</b>	Particional	Possui um cluster central os demais ligados a ele (KOWALSKI, 2007).
<b>Clique</b>	Particional	Adicionar um elemento a um determinado grupo, se este for similar a todos os demais. Este algoritmo busca a coesão entre os elementos do mesmo grupo (AGRAWAL et al., 1998).

A ideia de resultado ideal de clusterização varia entre diferentes aplicações e pode ser até diferente entre os usuários. Uma aplicação pode exercer influência no resultado de um algoritmo de clusterização, utilizando apenas subgrupos de atributos ou adaptando o uso de medidas de similaridade, e assim controlar o processo de clusterização.

Dentre os algoritmos listados no Quadro 3-2, é destacado o CLIQUE (AGRAWAL et al., 1998). Este algoritmo foi implementado devido as necessidades apresentadas pela perspectiva da mineração de dados:

- Tratamento efetivo de alta dimensionalidade: um objeto (registro de banco de dados) normalmente possui dúzias de atributos. Para Berchtold et al. (BERCHTOLD et al., 1997), não há significado na busca por grupos em que a média de encontro da dimensão dos dados é baixa. Muitas dimensões ou combinações de dimensões podem criar dados que são uniformemente distribuídos. Portanto, funções de distância que utilizam todas as dimensões de dados podem não ser efetivas. Além disso, muitos clusters podem existir em diferentes subespaços compostos de diferentes combinações de atributos (AGRAWAL et al., 1998);
- Interpretabilidade dos resultados: aplicações de mineração de dados normalmente necessitam de grupos que possam ser facilmente assimilados pelo usuário final (FAYYAD et al., 1996);

Segundo Kowalski (KOWALSKI, 2007), este algoritmo produz resultados idênticos independente da ordem nos quais os dados foram inseridos, pois, com a utilização deste algoritmo torna-se evidente a busca por uma coesão entre os elementos do mesmo grupo.

O algoritmo de Cliques pode ter seu comportamento descrito brevemente nos seguintes passos:

1. Seleciona o 1º elemento e coloca em um novo cluster;
2. Procura o próximo objeto similar;
3. Se o objeto é similar a todos os outros elementos do cluster, este objeto é agrupado;
4. Enquanto houver objetos, retornar ao passo 2;

5. Para os elementos não alocados (sementes de cluster), repetir o passo 1.

A Figura 3-1 apresenta o grafo do algoritmo Cliques.

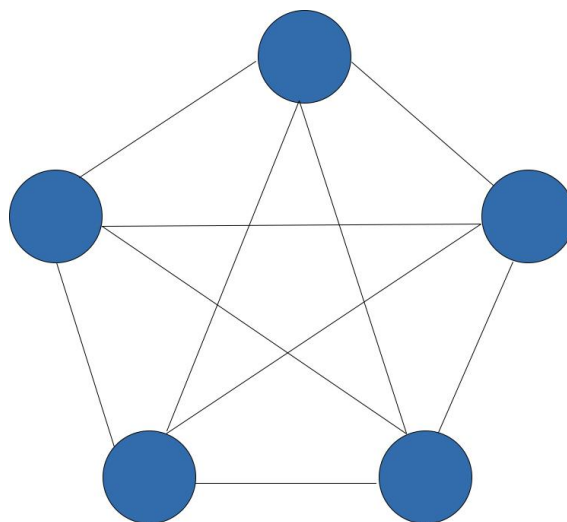


Figura 3-1. Grafo do Algoritmo Cliques (KOWALSKI, 2007).

### 3.3.3. Medidas de Similaridade

Apesar dos benefícios que os algoritmos de clusterização trazem para análises e utilização de dados para futuras recomendações, clusterizações precisas necessitam de definições de proximidade entre os pares de objetos, que são definidas como similaridades ou distâncias (HUANG, 2008).

Medidas de similaridade são métricas que permitem avaliar quantitativamente a semelhança entre dois objetos. Uma variedade de diferentes medidas de similaridade pode ser utilizada para calcular a similaridade entre dois itens. Nos últimos anos, várias medidas de similaridade foram propostas e largamente aplicadas, como a similaridade de Cosine e o coeficiente de similaridade de Jaccard (ERTOZ; STEINBACH; KUMAR, 2001). Medidas como a distância Euclidiana têm sido aplicadas em clusterização para calcular a distância de pares de objetos.

Em geral, medidas de similaridade/distância realizam um mapeamento entre a descrição simbólica de dois objetos e um único valor numérico. Uma característica de uma fórmula de similaridade é que os resultados são maiores

quando os itens são mais similares. O resultado é zero se os itens são totalmente dissimilares.

Dentre as medidas de similaridade/distância mais utilizadas, é destacado o Coeficiente de Similaridade de Jaccard, o qual mede a similaridade como a intersecção dividida pela união de objetos. Para documentos textuais, o coeficiente de Jaccard compara o peso da soma de termos compartilhados com o peso da soma de termos que estão presentes em ambos documentos, mas não são termos compartilhados.

Chapman (CHAPMAN, 2009) afirma que este algoritmo é baseado em similaridade de vetores. A cada ocorrência coincidente nos dois vetores V1 e V2, é incrementado o contador que irá gerar o grau de similaridade. O conteúdo dos vetores é separado em três conjuntos diferentes:

- A – que contém somente os valores coincidentes entre V1 e V2;
- B – que contém somente os valores não coincidentes de V1;
- C – que contém somente os valores não coincidentes de V2;

Então o coeficiente é calculado pela seguinte representação formal:

$$\text{Jaccard} = A / (A+B+C) \quad (1)$$

O domínio de similaridade do Coeficiente de Jaccard está entre 0 e 1. Isto é, quando os dois vetores V1 e V2 são idênticos, o coeficiente é 1, e 0 quando os vetores V1 e V2 são totalmente dissimilares.

Medidas de similaridade são largamente utilizadas em métodos de mineração de dados. Agrupar informações de acordo com a proximidade calculada, por meio de uma função de similaridade, para que posteriormente sejam apresentadas ao usuário de acordo com o seu perfil é uma tarefa que envolve o uso de medidas de similaridade em sistemas de recomendação, que serão tratados na Seção 3.4.

### **3.4 - Sistemas de recomendação**

Os principais objetivos dos sistemas de recomendação são a fidelidade e o conseqüente aumento da lucratividade de empresas. Diferentes estratégias podem ser adotadas para personalizar informações para um usuário, cada uma exigindo um grau de complexidade distinto no tratamento de informações coletadas (REATEGUI, CAZELLA, 2005).

Os Sistemas de Recomendação auxiliam no aumento da capacidade e eficácia deste processo de indicação já bastante conhecido na relação social entre seres humanos (RESNICK, VARIAN, 1997). Em um sistema típico, os usuários fornecem recomendações como entradas que o sistema agrega e direciona para os indivíduos considerados potenciais interessados neste tipo de recomendação. Um dos grandes desafios deste tipo de sistema é realizar a combinação adequada entre as expectativas dos usuários e os resultados recomendados.

Sistemas de recomendação utilizam um número variado de tecnologias. Estas tecnologias podem ser classificadas entre dois amplos grupos: Sistemas de Recomendação baseados em conteúdo; e Sistemas de Recomendação baseados em Filtragem Colaborativa;

#### Sistemas Baseados em conteúdo

Sistemas Baseados em conteúdo, examinam as propriedades dos itens selecionados em um primeiro momento pelo usuário, e posteriormente aplicam a recomendação de outros itens com base nas características iniciais preferidas pelo usuário. A Figura 3-2 apresenta o processo de recomendação baseada no conteúdo.



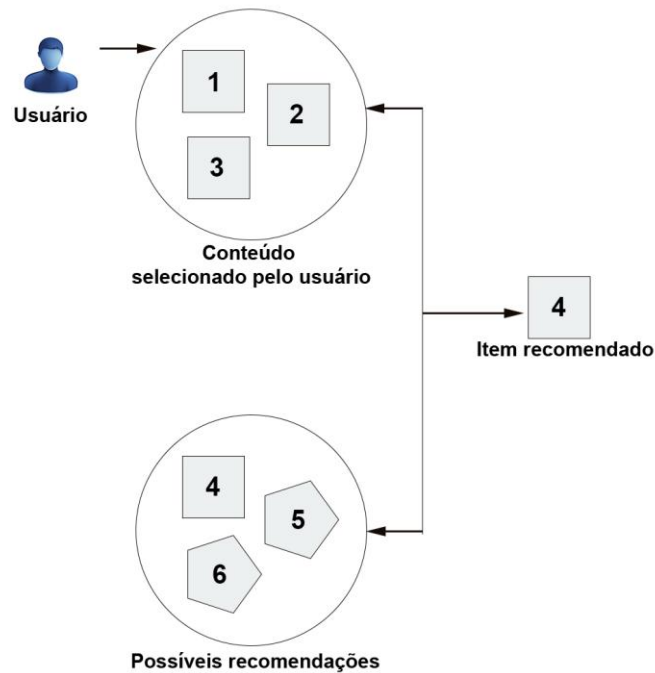
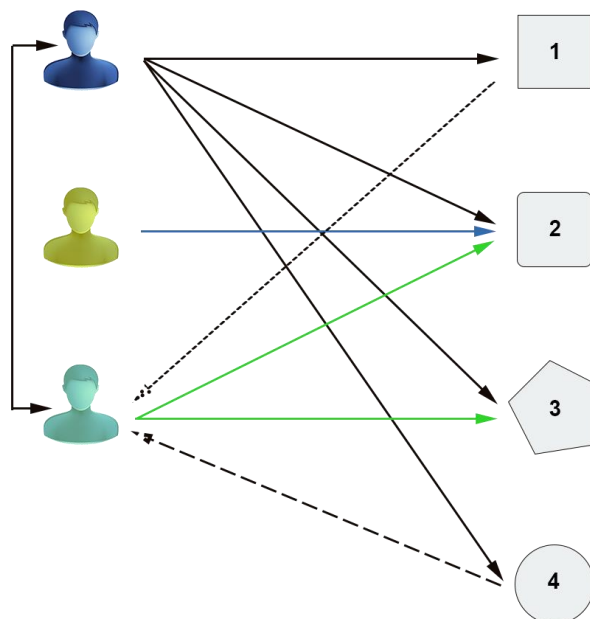


Figura 3-2. Processo de Recomendação Baseada em Conteúdo (LOPS; GEMMIS; SEMERARO, 2011).

### Sistemas de Recomendação baseado em Filtragem Colaborativa

Sistemas de Filtragem Colaborativa recomendam itens baseados em medidas de similaridade entre os usuários e os itens. Os itens recomendados a um usuário são aqueles preferidos por outros usuários similares. Este tipo de sistema de recomendação pode ser intensamente baseado nos temas das seções anteriores sobre Similaridade e Clustering. A Figura 3-3 apresenta o modelo de recomendação baseada em filtragem colaborativa de usuários. O sistema verifica os itens preferidos pelos usuários e então identifica os usuários que são similares por meios de medidas de similaridade e agrupamentos. Assim, o sistema pode apontar para possíveis recomendações que sejam de interesse de outros usuários.



**Figura 3-3. Processo de Recomendação Baseada no Usuário (LESKOVEC; RAJARAMAN; ULLMAN, 2014).**

Os autores Goldberg et al. (GOLDBERG et al., 1992) e Resnick e Varian (RESNIK; VARIAN, 1997), proponentes do primeiro Sistema de Recomendação, propuseram um tipo de recomendação denominado “filtragem colaborativa”, pois visava designar um processo de filtragem de informação com auxílio humano, ou seja, colaboração entre os grupos interessados. A expressão Sistemas de Recomendação é apresentada neste sistema por ser um termo genérico. Os autores defendem este posicionamento por dois motivos: primeiramente, pois os recomendadores podem não explicitar colaboração com os que as recebem, pois um pode não conhecer o outro, e por último os recomendadores podem sugerir itens de interesse particular, incluindo aqueles que poderiam ser desconsiderados.

Alguns autores, como Montaner, López e De La Rosa (MONTANER; LÓPEZ; DE LA ROSA, 2003), destacam que existe um terceiro tipo de recomendação de informação denominada de filtragem demográfica. A filtragem demográfica utiliza a descrição de um indivíduo, para aprender o relacionamento entre um item em particular e o tipo de indivíduo que poderia se interessar. O perfil do usuário é criado pela classificação dos usuários em estereótipos, que representam as características de uma classe de usuários.

Quanto às técnicas de filtragem de informação, além da filtragem colaborativa, que pode ser baseada em conteúdo ou no perfil do usuário, Burke (BURKE, 2002) define duas outras técnicas. Na primeira delas, intitulada de filtragem baseada em conhecimento, a recomendação dos itens é feita baseada nas inferências das preferências do usuário e suas necessidades, por meio de conhecimento estruturado de forma funcional. Alguns exemplos clássicos desta abordagem são: Google (BRIN; PAGE, 1998) e The Entree (Burke, 2002). Na segunda técnica, intitulada de filtragem baseada em utilidade, a recomendação é realizada considerando a utilidade dos itens para um determinado usuário.

González, Rosa e Montaner (GONZÁLEZ; ROSA; MONTANER, 2007) e Nunes (NUNES, 2009) propõem técnicas de recomendação complementares seguindo a linha proposta por Burke (BURKE, 2002). Essas técnicas envolvem questões psicológicas do usuário combinando as características dos serviços, e intitula-se filtragem baseada em outros contextos.

Segundo Adomavicius e Tuhilin (ADOMAVICIUS; TUZHILIN, 2005) e Burke (BURKE, 2002) as técnicas de filtragem são utilizadas com objetivo de melhorar a resposta da recomendação. Desta forma, a utilização de técnicas de filtragem são combinadas com técnicas de sistemas de recomendação. Esse tipo de combinação é denominado de filtragem híbrida.

### **3.5 - Descoberta de Conhecimento e Aprendizagem Organizacional**

Os documentos criados em uma organização representam o seu conhecimento em potencial. Potencial porque apenas parte de seus dados e informações serão consideradas úteis para serem utilizados em prol da aprendizagem organizacional (DAVANZO et al., 2008).

Neste ponto, um desafio encontrado é o de fazer a seleção de informações relevantes a partir de grandes quantidades de documentos, e a habilidade de disponibilizá-las para uso e reúso pelos membros da organização. Para auxiliar a enfrentar este desafio, a mineração de Textos, as técnicas de Sistemas de Recomendação, e o Processamento de Linguagem

Natural, podem ser utilizadas. Além disso, estas técnicas podem auxiliar no suporte efetivo para a Aprendizagem Organizacional, no quesito da extração de informação relevante de grandes quantidades de dados textuais não estruturados, e assim, criar o conhecimento (BORDONI et al., 2002).

No decorrer dos últimos 10 anos foram propostos vários trabalhos relacionando às áreas de recomendação de sistemas e mineração de textos. Dentre estes trabalhos, estão pesquisas como a de Reichling, Veith e Wolf (REICHLING; VEITH; WOLF, 2007), que propõe um sistema de recomendação de *expertise* para as necessidades específicas de uma associação industrial nacional europeia. Outros estudos têm propostos sistemas para gerenciamento de conhecimento, como o de Luo e Cao (LUO; CAO, 2009) que apresenta uma arquitetura para realizar o compartilhamento de conhecimento e recomendação de conhecimento baseado em modelo de usuário. Também é apresentado por Ale et al., (ALE et al., 2014), uma arquitetura de sistema capaz de fornecer suporte tecnológico para representação de conteúdos e atividades de recuperação de conhecimento.

Estes trabalhos apontam que continuamente, as pesquisas na área de aprendizagem organizacional têm buscado novas tecnologias, podendo ser aplicadas em diferentes domínios do conhecimento.

### **3.6 - Descoberta de Conhecimento para Engenharia de Software**

Desenvolvedores de software são continuamente apresentados a novas tecnologias, componentes e ideias, enquanto que os sistemas nos quais trabalham tendem a crescer e a depender de uma lista crescente de bibliotecas externas e recursos. Desenvolvedores de software devem, como rotina, gastar uma larga fração de seu tempo de trabalho procurando por informação, por exemplo, para entender um código existente ou descobrir como implementar um código de forma apropriada (EKOKO; ROBILLARD, 2012). Um dos maiores desafios para os engenheiros de software, na execução de uma tarefa de manutenção, por exemplo, é a necessidade de compreender uma variedade de artefatos, que estão frequentemente desconectados, criados originalmente como parte do processo de desenvolvimento de software (JIN; CORDY, 2005).

Estes artefatos incluem, entre outros, código fonte, especificações de requisitos, interfaces e guias de usuários. Da perspectiva de um desenvolvedor, é essencial estabelecer e manter conexões semânticas entre todos estes artefatos. No entanto, integrar a grande quantidade de documentação requer novas abordagens, que vão desde a análise de documentos em linguagem natural (PLN), à evoluções de técnicas de recuperação de informação, minerações de dados e recomendações (ANTONIOL et al., 2000).

Afirmações como essa, trouxeram interesse e empenho para associação entre tecnologias de descoberta de conhecimento e a engenharia de software, que resultaram em trabalhos como o de Li e Rilling (LI; RILLING, 2008), que apresenta uma abordagem para analisar documentações de software em um nível semântico. Robillard e Walker (ROBILLARD; WALKER, 2014) definem dentro da engenharia de software, uma série de áreas nas quais os sistemas de recomendação podem atuar, como:

- Código fonte de um projeto: sistemas de recomendação podem ajudar desenvolvedores a navegar pelo código fonte do seu próprio projeto, por exemplo, tentando encontrar áreas do código fonte nas quais o desenvolvedor precise;
- Código fonte reusável: outros sistemas de recomendação em engenharia de software auxiliam usuários a descobrir os elementos de uma API (*Application Programming Interface*) (como classes, funções, scripts) que possam auxiliar a completar uma tarefa;
- Relatórios de problemas: a maioria do conhecimento sobre um projeto de software pode estar em sua base de problemas. Quando um desenvolvedor está empenhado em resolver um problema, sistemas de recomendação podem auxiliar a descobrir problemas relacionados que já foram reportados;
- Pessoas: em algumas situações os sistemas de recomendação podem também ajudar a encontrar a melhor pessoa para resolver uma tarefa, ou o especialista para auxiliar na implementação da solução.

### **3.7 - Considerações Finais**

Este capítulo destacou os motores de busca como uma importante forma de coleta de necessidades em construção de competências. Pelo fato de motores de buscas serem os principais caminhos pelos quais os desenvolvedores de software encontram respostas para problemas que surgem no dia a dia em projetos de desenvolvimento.

Também foram apresentados conceitos sobre a área de mineração de textos, bem como as principais técnicas de processamento e clusterização de textos. Uma breve apresentação da área de sistemas de recomendação e da importância da integração destas tecnologias para a evolução do campo da aprendizagem organizacional e também da engenharia de software foi realizada.

## Capítulo 4 - **ESTRUTURAÇÃO DA PESQUISA**

Este capítulo apresenta alguns conceitos importantes referentes à metodologia e métodos de pesquisa, de forma a posicionar o leitor quanto às opções metodológicas adotadas nesta pesquisa e detalha as estratégias de pesquisa para atingir o objetivo geral do trabalho.

### **4.1 - Estruturação da Pesquisa**

Este trabalho, por se tratar da continuação de um estudo apresentado em (MENOLLI, 2013), adota a mesma abordagem de pesquisa caracterizada em seu trabalho, combinando vários métodos e procedimentos técnicos, como pode ser visto na Figura 4-1.

A estrutura é dividida em quatro fases (Planejamento Inicial, Fase exploratória, Desenvolvimento, Avaliação e Conclusão), sendo que cada fase contém uma ou mais etapas. Estas etapas são relacionadas entre si. Os relacionamentos indicados com setas contínuas, indicam a etapa base para definição da etapa corrente. Os relacionamentos indicados com setas pontilhadas indicam os processos que são influenciados pela etapa corrente. As fases da estrutura desta pesquisa são descritas a seguir.

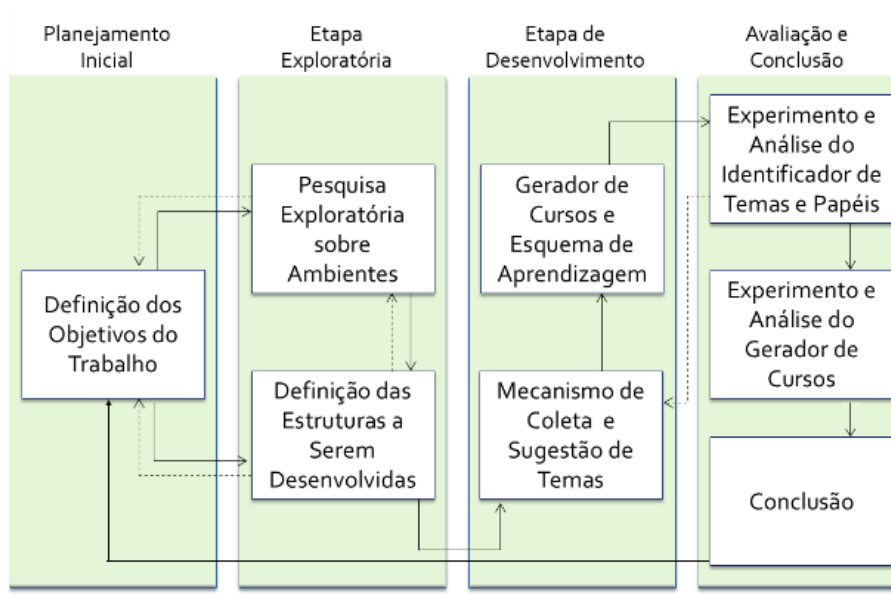


Figura 4-1. Estrutura da Pesquisa

#### 4.1.1. Planejamento Inicial

Após a apresentação das lacunas existentes no ambiente proposto em por Menolli, Malucelli e Reinehr (MENOLLI; MALUCELLI; REINEHR, 2011), foi definido o objetivo geral deste trabalho que é **desenvolver um mecanismo para criação semiautomática do esquema de aprendizagem em um ambiente colaborativo semântico voltado à aprendizagem organizacional para empresas de desenvolvimento de software.**

Juntamente com o objetivo geral do trabalho foram definidos os objetivos específicos, com intuito de direcionar a pesquisa em suas fases de desenvolvimento.

O primeiro objetivo específico, que é desenvolver um mecanismo para captar as consultas em motores de busca, refere-se principalmente a auxiliar na limitação presente na abordagem de Menolli, Malucelli e Reinehr (MENOLLI; MALUCELLI; REINEHR, 2011), citada na introdução deste trabalho, que é o alto grau de esforço envolvido na tarefa de coleta das necessidades de aprendizagem da equipe de desenvolvimento de software.

O segundo objetivo específico, que propõe uma abordagem para a criação semiautomática do esquema de aprendizagem, está relacionado especificamente com a dificuldade em descrever o arquivo em metalinguagem



(XML) seja por um especialista ou qualquer outro membro da equipe de desenvolvimento.

Por fim, o terceiro objetivo específico, que é avaliar o mecanismo proposto, tem como foco validar as abordagens propostas, por meio de experimento em ambiente controlado, e posteriormente analisar e apresentar os resultados obtidos na avaliação de forma descrita e qualitativa.

#### **4.1.2. Fase Exploratória**

A fase exploratória foi subdividida em duas etapas:

- primeira etapa: realização de uma pesquisa exploratória, com foco em atualizar o cenário da aprendizagem organizacional com relação às abordagens existentes voltados a aprendizagem organizacional e; e
- segunda etapa: definição das estruturas a serem desenvolvidas neste trabalho;

##### Pesquisa exploratória sobre ambientes

A fase exploratória tem como objetivo proporcionar maior familiaridade com o problema com vistas a torná-lo explícito (GIL, 2002). Assim, é necessário explorar o objeto de estudo, de forma a definir o que, dentro da criação do curso, continua sendo tarefa de um especialista, e o que será semiautomático, uma vez que o foco deste estudo não é tornar a criação do esquema de aprendizado totalmente automática.

Nesta fase, uma vertente foi seguida para compreender o campo de estudo. A partir de fontes de dados na forma bibliográfica uma revisão da literatura foi conduzida de forma exploratória. O levantamento bibliográfico se faz necessário uma vez que a revisão da literatura é o meio pelo qual o pesquisador pode identificar o conhecimento científico existente em uma determinada área, de forma a planejar sua pesquisa, evitando a duplicação de esforços e a repetição de erros passados (MAFRA; TRAVASSOS, 2006). Entretanto, a revisão da literatura deve ser conduzida de uma forma confiável e abrangente, para que seus resultados tenham valor científico. A revisão de

literatura deve ser conduzida com um protocolo de revisão pré-estabelecido de maneira a minimizar os interesses pessoais de seus pesquisadores, levando a resultados confiáveis (MAFRA; TRAVASSOS, 2006).

Assim, definiu-se por realizar uma pesquisa exploratória utilizando-se os mesmos termos de pesquisa apresentados na revisão sistemática realizada em Menolli, Reinehr e Malucelli (MENOLLI; REINEHR; MALUCELLI, 2013a). Estes termos são apresentados no Quadro 4-11 na coluna “Termos de Engenharia de Software e Aprendizagem Organizacional”. Esta pesquisa foi conduzida com objetivo de atualizar o universo de trabalhos de engenharia de software com foco na aprendizagem organizacional.

Entretanto, como a abordagem proposta nesta dissertação, para coleta de temas e papéis, e criação do esquema de aprendizagem, se faz por utilização de técnicas de recuperação de informação, mineração de texto e clusterização, foram combinados aos temas de pesquisa, termos referentes a estas tecnologias. Estes termos também são apresentados no Quadro 4-1 na coluna “Termos de Sistemas de Recomendação e Mineração de Texto”.

**Quadro 4-1. Termos da Pesquisa (MENOLL; REINEHR; MALUCELLI, 2013a).**

Termos de Engenharia de Software e Aprendizagem Organizacional	Termos de Sistemas de Recomendação, Mineração de Texto
Software Engineering Software Process Software Learning Software Organization Learning Organization Organizational Learning Knowledge Management	Web Information Retrieval Recommendation Systems Clustering Algorithm for Text Mining Similarity

#### Definir as Estruturas a Serem Desenvolvidas

A pesquisa exploratória proveu uma visão geral do campo de estudo de ambientes voltados à aprendizagem organizacional que utilizam técnicas de mineração de textos e sistemas de recomendação. Além disso, os objetivos iniciais delimitaram o escopo do trabalho, auxiliando a identificar o ambiente a ser desenvolvido.

A etapa exploratória auxiliou na definição da seguinte abordagem que ocorre em duas etapas. A primeira etapa é a implementação de um mecanismo coletor de consultas realizadas por desenvolvedores de software em motores de busca, que processa as consultas e classifica-as de forma a sugerir ao especialista os temas de cursos que representem a necessidade de aprendizagem da equipe de desenvolvimento. A segunda etapa é o desenvolvimento de um ambiente para que o especialista possa localizar os temas relevantes por meio do mecanismo de sugestão de temas, e com base em uma abordagem baseada em conceitos educacionais de elaboração de cursos, este especialista poder criar o Esquema de Aprendizagem.

Estas duas etapas foram implementadas e serão apresentadas no Capítulo 5.

#### **4.1.3. Desenvolvimento**

Esta fase foi dividida em duas partes: mecanismo de Coleta e Sugestão de Temas, e Gerador de Cursos e Esquema de Aprendizagem, que são complementares e dependentes entre si, descritas a seguir.

##### Mecanismo de Coleta e Sugestão de Temas

Primeiro foi implementada a extração de informações das pesquisas realizadas pelos desenvolvedores em motores de busca utilizando técnicas de mineração de textos como por exemplo, tokenização, filtragem, stemming, e baseado em sistemas de recomendação.

Além da pesquisa realizada pelo usuário, por meio do motor de busca, faz-se necessário obter o papel que este usuário possui dentro da empresa, de modo que seja possível nivelar o curso que será gerado, atendendo ao nível de capacidade do usuário. Para tal, é identificar o conhecimento dos usuários dentro da organização. Esta identificação trata-se de um cadastro no qual o colaborador será inserido juntamente com suas habilidades e o nível que possui em cada uma.

Posteriormente, foi desenvolvido um mecanismo que, por meio das informações coletadas pelas buscas dos usuários na *web*, possa sugerir ao especialista gerador de unidades quais temas estão sendo pesquisados, e qual

o papel desempenhado por estes usuários. Com isto, o especialista poderá escolher o tema para definir o curso. Para que as informações como a seleção dos participantes, o nível, a hierarquia das atividades do curso, bem como os objetivos, fossem selecionadas, foi implementado um ambiente que é dividido em 4 passos para criação do curso. No primeiro passo o especialista seleciona informações gerais do curso, como objetivos e requisitos. No segundo passo, o especialista informará temas adicionais. No terceiro passo, os temas poderão ser agrupados de forma hierárquica, de forma a organizar como o curso será apresentado. Por fim, a última etapa permite que o especialista selecione os participantes e crie o Esquema de Aprendizagem.

#### Gerador de Cursos e Esquema de Aprendizagem

A próxima etapa consistiu em desenvolver um ambiente capaz de definir o esquema de aprendizagem, criando uma estrutura que inicialmente foi definida pelo Esquema de Aprendizagem. Este ambiente foi baseado em conceitos educacionais e em teorias de elaboração de cursos, assim, o especialista, no momento da criação do curso, pode definir cursos que sejam efetivos para o aprendizado.

Por fim, foi desenvolvido um componente para gerar o Esquema de Aprendizagem, que no ambiente colaborativo semântico é responsabilidade do especialista, para poder seguir o processo de geração de cursos.

Após as etapas anteriores terem sido finalizadas, um ambiente incluindo o mecanismo proposto foi implementado em linguagem de programação C#, (C-SHARP).

#### **4.1.4. Avaliação e Análise**

A última fase da estrutura da pesquisa apresentada na Figura 4-1 é a de avaliação e conclusão do trabalho. Esta fase foi dividida em três etapas. As duas primeiras são referentes a avaliação do ambiente implementado. Por fim, após a avaliação do ambiente, a última etapa, de conclusões do trabalho, é executada.

### Experimento e análise do identificador de temas e papéis

Inicialmente foi necessário realizar o planejamento para condução do estudo para avaliar a abordagem proposta neste trabalho. Foi definido utilizar um estudo experimental, que tem como objetivo colher dados, em um ambiente controlado, para confirmar ou negar uma hipótese. O método experimental considera a proposição e avaliação do modelo com os estudos experimentais (TRAVASSOS; GUROV; AMARAL, 2002).

No entanto, Travassos, Gurov e Amaral (TRAVASSOS; GUROV; AMARAL, 2002) ressaltam que:

“É importante notar que os experimentos não provam nada. É verdade que nenhum experimento oferece prova com certeza absoluta. Os experimentos verificam a previsão teórica de encontro à realidade. A comunidade aceita uma teoria se todos os fatos conhecidos dentro de seu domínio possam ser deduzidos da teoria, possuía uma verificação experimental extensa e prediz o novo fenômeno corretamente (p.4).”

Com o intuito de planejar os experimentos, foram primeiramente analisados os modelos de experimentação aplicáveis para novas tecnologias de software apresentado por Zelkowitz, Wallace e Binkley (ZELKOWITZ; WALLACE; BINKLEY, 2003). Neste trabalho é exposta uma classificação a partir dos métodos de validação dos experimentos, e estes são divididos em cinco categorias: Observação, Histórico, Controlado, Formal e Informal. Cada uma destas categorias é descrita abaixo (ZELKOWITZ; WALLACE; BINKLEY, 2003):

1. Observação: estes métodos envolvem o monitoramento do projeto, como ele se desenvolve, de forma a coletar dados sobre a eficácia de uma nova tecnologia.
2. Históricos: estes métodos envolvem uma análise de dados coletados com o intuito de descobrir o que aconteceu durante o desenvolvimento de um projeto desenvolvido anteriormente.
3. Controlados: estes métodos envolvem estudo cuidados de estratégias alternativas para determinar a eficácia de um método

em comparação com outros métodos. Este é o conceito mais tradicional quando se pensa em um “experimento”.

4. Formais: esses métodos envolvem o uso de um modelo formal para descrever um processo. A validação final depende de outro método para determinar se o modelo formal concorda com a realidade.
5. Informais: são geralmente ad hoc e não fornecem resultados significativos.

Dentro das cinco categorias, Zelkowitz, Wallace e Binkley (ZELKOWITZ; WALLACE; BINKLEY, 2003) propuseram uma classificação de quatorze métodos de validação. No Quadro 4-2, é apresentado resumidamente cada um destes métodos, assim como os seus principais pontos fortes e fracos.

Considerando que foi definido utilizar estudos experimentais, optou-se por utilizar um método controlado para validação destes estudos. Dentre os métodos controlados apresentados no Quadro 4-2, optou-se pelo de simulação. Neste método o pesquisador executa o produto com dados artificiais. No desenvolvimento de software, os projetos são geralmente grandes, contendo pessoas de vários projetos. Este cenário em um ambiente real é proibitivamente caro. Por este motivo, a maioria das repetições de engenharia de software é executada em um ambiente artificial pequeno (ZELKOWITZ; WALLACE; BINKLEY, 2003). Isto fornece percepções sobre a eficácia do método, no entanto uma vez que o método é aplicado isoladamente, o impacto deste método em relação aos outros métodos utilizados nos projetos não é imediatamente aparente (ZELKOWITZ; WALLACE; BINKLEY, 2003).

Após o método ter sido definido, o experimento foi planejado de acordo com seu objetivo.

Este experimento tem como objetivo avaliar o mecanismo que coleta as buscas dos desenvolvedores em motores de busca e sugere temas para criação de cursos. Neste experimento foram formuladas hipóteses que puderam validar a abordagem proposta.

**Quadro 4-2. Métodos de Validação para Experimentação. Adaptado de (ZELKOWITZ; WALLACE; BINKLEY, 2003)**

<b>Categoria</b>	<b>Método de Validação</b>	<b>Descrição do Método</b>	<b>Pontos Negativos</b>	<b>Pontos Positivos</b>
<b>Observação</b>	Monitoramento de Projeto	Coleta de dados de desenvolvimento	Não especifica objetivos	Fornece base para o futuro;
	Estudo de Caso	Monitora um projeto em profundidade	Controles pobres para replicações posteriores	Consegue analisar fatores restritos a baixo custo
	Estudo de campo	Monitora múltiplos projetos	Tratamentos diferem entre projetos	Forma barata de replicação
<b>Histórico</b>	Pesquisa bibliográfica	Examina estudos anteriores publicados	Viés de seleção; tratamentos diferem	Grandes bases de dados disponíveis; barato
	Dados legados	Examina dados de projetos concluídos	Não pode restringir fatores; dados limitados	Combina múltiplos estudos; barato
	Lições aprendidas	Examina os dados qualitativos a partir de projetos concluídos	Não há dados quantitativos, não pode restringir fatores	Determina tendências; barato
<b>Controlado</b>	Replicação	Desenvolve várias versões do produto	Muito Caro	Pode controlar fatores de risco para todos os tratamentos
	Sintético	Replica um fator em ambiente de laboratório	Interações entre múltiplos fatores	Pode controlar fatores individuais; custos moderados
<b>Formal</b>	Análise Teórica	Uso da lógica formal para provar o valor da tecnologia	Não fica claro se o modelo formal concorda com a realidade	Validação barata; Se o modelo está correto, o método é eficaz
<b>Informal</b>	Assertion	Validação Ad hoc	Validação Insuficiente	Base para futuros experimentos

Após o método e o objetivo terem sido definidos, foi escolhido o domínio a ser aplicado o experimento. Foi selecionado um domínio de desenvolvimento de software, relacionado aplicações de sistemas empresariais.

Após definido o domínio, o experimento foi concebido e executado, tendo como população quatro colaboradores de uma empresa de desenvolvimento de software. A análise dos dados foi realizada de forma qualitativa, por meio da análise textual discursiva (MORAES; GALIAZZI, 2006).

Este experimento foi denominado experimento para avaliar a identificação de temas e foi seus detalhes são apresentados no Capítulo 6;

#### Experimento e análise do Definidor de Cursos

Após a execução do primeiro experimento, um segundo experimento foi realizado. Este segundo experimento também foi realizado por meio de estudo experimental, e foi planejado com base no primeiro, assim utilizou o mesmo domínio, e tendo como participantes dezoito colaboradores de organizações diferentes.

O objetivo deste experimento foi avaliar o ambiente desenvolvido para criação de cursos. Além de colaboradores de empresas de desenvolvimento de software, dois especialistas que foram denominados de especialista técnico e especialista de recursos humanos, foram consultados e participaram do experimento. O especialista técnico foi responsável pela criação de um curso e do esquema de aprendizagem utilizando o ambiente implementado, e o especialista de recursos humanos atuou na análise do ambiente e do curso criado.

Para avaliar o ambiente, este experimento foi executado e os resultados foram analisados usando um método qualitativo, adaptando a metodologia de análise de conteúdo, proposta por Bardin (BARDIN, 1977). Este experimento foi denominado experimento para avaliar o ambiente, e é apresentado em detalhes no Capítulo 7.

#### Conclusões do Trabalho

Por fim, após o planejamento, execução e avaliação dos dois experimentos, foi possível concluir o trabalho. Utilizando a análise dos dados dos dois experimentos executados, em conjunto com o referencial teórico obtido no decorrer do projeto, as conclusões finais do trabalho foram expostas.



## **4.2 - Considerações Finais**

Este capítulo apresentou a estruturação da pesquisa, bem como a estratégia de pesquisa que definem as etapas que conduziram a pesquisa e as fases a serem seguidas neste estudo.

## Capítulo 5 - O AMBIENTE GERADOR DE ESQUEMA DE APRENDIZAGEM

Este capítulo tem como objetivo apresentar detalhadamente o mecanismo para a geração semiautomática do esquema de aprendizagem. O propósito de semiautomatizar a criação do Esquema de Aprendizagem é de posteriormente localizar objetos de aprendizagem em um ambiente colaborativo semântico voltado para aprendizagem organizacional. Por esse motivo, neste capítulo será apresentada a abordagem desenvolvida, a arquitetura delineada para o funcionamento do mecanismo coletor de temas, assim como as características visuais e funcionais do ambiente proposto.

### 5.1 - Introdução

Por este trabalho se tratar de uma abordagem proposta para suprir algumas limitações identificadas no ambiente colaborativo semântico proposto por Menolli, Malucelli e Reinehr (MENOLLI; MALUCELLI; REINEHR, 2011), se faz necessário definir claramente qual a função do mecanismo proposto neste trabalho, quando incorporado ao ambiente colaborativo semântico.

O ambiente colaborativo semântico proposto por Menolli, Malucelli e Reinehr (2011) tem como objetivo principal organizar o conteúdo gerado em forma de objetos de aprendizagem, e então, utilizar um Esquema de Aprendizagem definido por um especialista, para criar unidades de aprendizagem. Como apresentado na Figura 5-1, o trabalho de definição do Esquema de Aprendizagem, exige alto nível de esforço por parte do especialista por se tratar de uma tarefa manual. O especialista é responsável por duas tarefas: (1) Definir o Esquema de Aprendizagem; e (2) Escolher Objetos que contemplarão uma unidade de aprendizagem. Portanto, o foco desta dissertação de mestrado está exatamente em auxiliar o especialista na definição do Esquema de Aprendizagem.

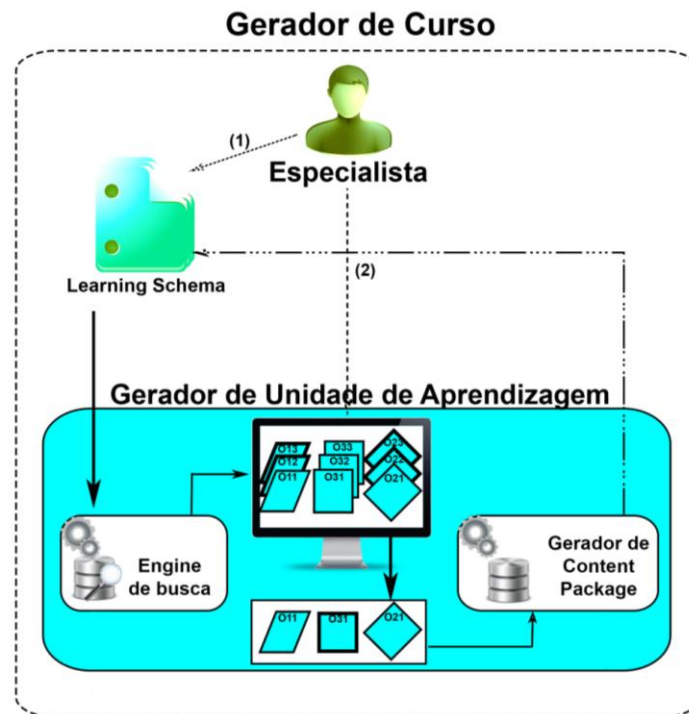


Figura 5-1. Gerador de Cursos (MENOLLI; REINEHR; MALUCELLI, 2013b)

Para isso, a proposta desta pesquisa é permitir que a partir de conteúdos inseridos por usuários em uma ferramenta social (*wiki*), sejam criados objetos e unidades de aprendizagem organizacionais semiautomaticamente, como apresentado no Capítulo 2. Nesta dissertação de mestrado é apresentada uma extensão do papel do especialista no ambiente colaborativo semântico, conforme Figura 5-2.

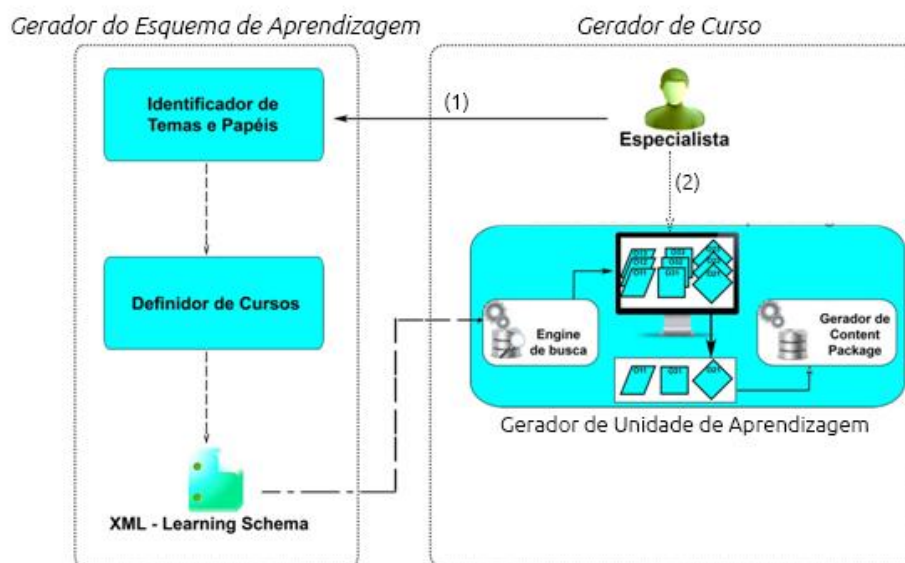


Figura 5-2. Extensão do Gerador do Esquema de Aprendizagem.

Como pode ser observado na Figura 5-1, o componente Gerador de Curso (MENOLLI; REINEHR; MALUCELLI, 2013b), apresenta dois passos executados pelo especialista: o primeiro passo é a criação do Esquema de Aprendizagem e o segundo passo é a escolha de objetos que contemplarão a unidade de aprendizagem. Na proposta, apresentada na Figura 5-2, que é uma extensão do Gerador de Cursos (MENOLLI; REINEHR; MALUCELLI, 2013b), foi incluída a etapa de criação do esquema de aprendizagem por meio do mecanismo Gerador de Esquema de Aprendizagem. Este mecanismo é dividido em dois elementos principais: Identificador de Temas e Papéis; e Definidor de Cursos.

O primeiro elemento, chamado de “Identificador de Temas e Papéis” (ITP), funciona como um coletor de consultas digitadas em motores de busca e informações sobre o usuário. Este componente utiliza as informações coletadas para sugerir ao especialista os temas e papéis mais pesquisados (TIRONI et al., 2015).

O segundo componente define um curso a partir de toda a informação coletada no ITP e define a estrutura da Unidade de Aprendizado. Este componente é chamado de “Definidor de Curso” (DC). Como tarefa final, baseado no IMS *Learning Design* (IMS, 2003), o DC irá criar o Esquema de

Aprendizagem, o qual será usado para criar a Unidade de Aprendizagem (TIRONI et al., 2015).

Neste novo processo, o especialista, ao invés de criar manualmente o Esquema de Aprendizagem, fará a escolha do tema do curso e a definição do curso, e o ambiente definidor de cursos fará a criação do Esquema de Aprendizagem. Depois deste processo, o Esquema de Aprendizagem poderá ser utilizado novamente no Gerador de Cursos (MENOLLI; REINEHR, MALUCELLI, 2013b).

As próximas seções apresentam de forma detalhada a concepção das abordagens propostas, bem como as especificações destas abordagens. Também é apresentado como o mecanismo gerador do Esquema de Aprendizagem foi implementado.

## **5.2 - Identificador de Temas e Papéis**

A abordagem proposta para o componente Identificador de Temas e Papéis tem como objetivo apresentar ao especialista quais temas estão sendo mais pesquisados pelos colaboradores, e qual o perfil do usuário que realiza esta pesquisa. Estes temas são originados a partir de dados coletados nos registros de *log* da rede empresarial, referentes às buscas realizadas pelos colaboradores em motores de busca, como por exemplo, Google, Yahoo e Bing.

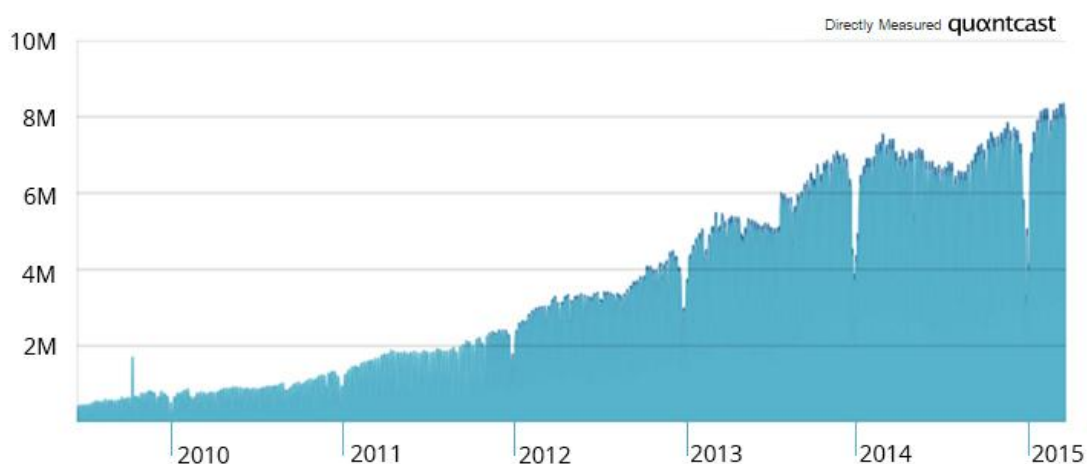
A maioria dos profissionais de desenvolvimento de software utiliza-se de serviços de fóruns de ajuda, pesquisas, e-mail corporativo e até mesmo ferramentas de mensagens instantâneas e redes sociais para troca de informações entre colegas de empresa ou ainda entre profissionais da mesma área de atuação (PETRI; OLIVEIRA; FONTOURA, 2013).

O uso e aplicação do conhecimento e propriedade intelectual coletiva, de forma voluntária, por meio de uma rede distribuída de indivíduos, para que resolvam problemas, criem soluções ou desenvolvam novas tecnologias, é denominado *crowdsourcing* (BRABHAM, 2008). Isto é, o uso do conhecimento coletivo, que pode ser construído por ferramentas de suporte *web*, fóruns *online* e até mesmo por meio de redes sociais, possibilitam um aumento da

produtividade do colaborador, que significa que a aprendizagem deste funcionário está sendo efetiva.

No entanto, a razão de escolher as consultas digitadas em motores de busca, como fonte de informação, se dá pelo fato de que mais e mais desenvolvedores de software estão utilizando motores de busca para encontrar técnicas, problemas de codificação resolvidos e novas tecnologias (MANNING, RAGHAVAN, SCHUTZE, 2008).

De acordo com o QuantCast, o fórum de programação “StackOverflow”, teve um crescimento de cerca de mil acessos para mais de três milhões de visitas por dia desde 2009 (QuantaCast, 2015). Este resultado, apresentado na Figura 5-3, mostra que houve um aumento considerável de programadores que passaram a capturar respostas para suas necessidades na internet, submetendo uma questão ou procurando repostas para uma questão que já foi realizada.



**Figura 5-3. Crescimento de Acessos ao Fórum Online StackOverflow (QuantaCast, 2015).**

Para este trabalho, foram consideradas diferentes abordagens que pudessem ser utilizadas para identificação dos temas relevantes para o aprendizado da equipe de desenvolvimento. Dentre elas, foi considerada como uma possível abordagem para coleta de temas neste trabalho, a criação de uma aplicação de navegador. Nesta aplicação de navegador, o usuário teria uma opção de ativar o arquivamento do histórico de buscas. Uma vez que o aplicativo estivesse conectado ao navegador, seriam utilizadas técnicas de *web*

*crawling* para visitar as páginas de fóruns acessadas pelo usuário, a fim de coletar os textos dos fóruns e assim descobrir, por meio do uso de técnicas de mineração de texto, assuntos relevantes para criação de cursos.

Também foi considerada como opção para coleta de dados e temas nesta dissertação, a criação de uma aplicação *online* na qual o usuário, ao invés de informar a sua consulta em um motor de busca, a consulta seria informada nesta aplicação. Esta aplicação seria responsável por enviar a busca do usuário para um motor de busca semântico para coletar os resultados. Uma vez que a consulta fosse informada em uma aplicação própria, o interesse do usuário seria facilmente coletado.

No entanto, estas abordagens quando aplicadas em um ambiente corporativo real, demandariam atenção dos desenvolvedores em informar suas dúvidas em uma ferramenta diferente de um motor de busca. Os motores de busca são os principais meio pelo qual os desenvolvedores recorrem às suas dúvidas (PRASATH et al., 2014). Por este motivo, estas abordagens se tornaram inviáveis, pois, envolveriam definições de processos organizacionais e políticas, que são assuntos fora do escopo de tratamento desta pesquisa.

Portanto, para este componente, que tem como principal objetivo apresentar sugestões de temas de cursos e usuários que podem participar ou lecionar estes cursos, optou-se por fazer coletas das consultas realizadas pelos colaboradores em motores de busca por meio dos registros do *log* de servidor *proxy*.

Após a coleta dos registros de *log*, é possível reunir o assunto de interesse do desenvolvedor, por exemplo, “Polimorfismo Exemplo Java”, e a informação do colaborador como o endereço IP (*Internet Protocol*), data e hora de busca. Então, são utilizadas práticas de mineração de texto e um algoritmo de clusterização para agrupar estes dados e descobrir qual foi o tema mais pesquisado pelo grupo.

Para obter estes temas é necessário delinear uma estratégia que, a partir de uma rede interna empresarial, seja possível coletar os registros de buscas. Uma das ferramentas mais usadas pelos colaboradores de TI para gerenciamento da rede interna, é o *Proxy*. Um *Proxy* é um servidor (um sistema de computador ou uma aplicação) que age de forma intermediária nas

solicitações de usuários que buscam fontes de informações em outros servidores (SHAPIRO, 1986).

Com o registro das solicitações de buscas dos usuários na internet, por meio de um Servidor *Proxy*, é possível que a partir de regras pré-especificadas, seja gerado um relatório destes registros. Estas regras funcionam como filtros, como por exemplo, de domínios de sites.

Dentre os *proxys* mais utilizados está o Squid. O Squid é um servidor *Proxy* que suporta os protocolos HTTP, HTTPS e FTP (SQUID, 2015). Os navegadores dos usuários da rede podem então utilizar o Squid local como um servidor *Proxy*. A principal utilização do servidor Squid em empresas de desenvolvimento de software é para o controle de acesso à internet. O Squid é capaz de armazenar registros sobre os acessos, incluindo as URLs acessadas, a data e a hora exatas, e qual colaborador dentro da rede realizou o acesso.

Assim, considerando estes conceitos, a primeira etapa para identificação de temas e papéis, que é realizar a coleta dos registros de busca de um *Log* de Servidor *Proxy*, é apresentada na Figura 5-4.

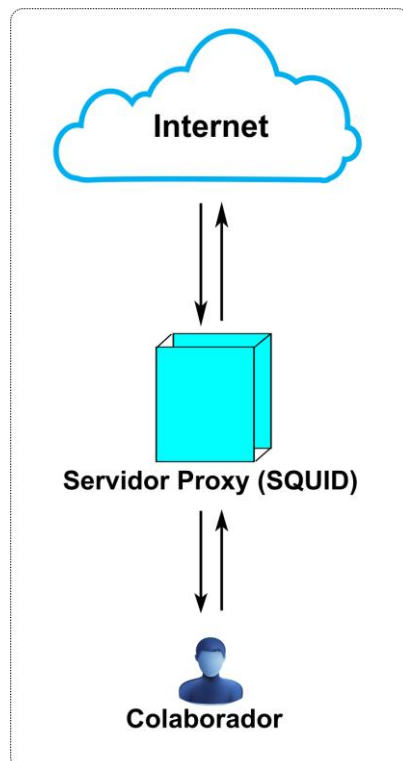


Figura 5-4. Servidor Proxy com SQUID.



Com um Servidor *Proxy* (Squid) mediando a conexão entre o usuário e a internet, é possível capturar as informações sobre cada consulta realizada no motor de busca (TIRONI et al., 2015). No entanto, um *log* de registros de acesso normalmente não é um arquivo de fácil leitura. A Figura 5-5 apresenta uma amostra de registros resultantes de uma consulta realizada no Google.

```
1432685442.952    325 192.168.1.101 TCP_MISS/200 957 GET http://www.google.com.br/
1432685443.221    487 192.168.1.101 TCP_MISS/200 955 GET http://www.google.com.br/
1432685443.258    313 192.168.1.101 TCP_MISS/200 959 GET http://www.google.com.br/
1432685462.611    428 192.168.1.101 TCP_MISS/200 923 GET http://www.google.com.br/
1432685462.780    357 192.168.1.101 TCP_MISS/200 926 GET http://www.google.com.br/
```

Figura 5-5. Amostra de registro do arquivo de *logs* de acesso ao Google.

Para coletar informação útil destes registros é necessário entender como funciona sua estruturação. Como pode ser verificado na Figura 5-5, estes registros são divididos em colunas. A primeira coluna refere-se à data e hora universal do sistema operacional Unix, que depois de convertida, se torna de fácil leitura, por exemplo:

1432685442.952 = Terça-feira, 26 de Maio de 2015 às 21:10:00

A segunda coluna representa o tempo (em milissegundos) de duração da solicitação. Este dado não é considerado significativo para a abordagem, por ser apenas uma informação técnica da rede.

A terceira coluna refere-se ao endereço IP do usuário que realizou a conexão. Empresas usualmente têm um endereço fixo para todos os colaboradores ou computadores. Dessa forma é possível indicar os funcionários que realizaram a conexão. O ambiente proposto permite que o especialista registre as informações dos colaboradores, como nome, habilidades e então, vincular esta informação com o endereço IP utilizado pelos funcionários dentro da rede da corporação.

Na quarta coluna é localizado o registro da ação tomada pelo *Proxy*. Neste campo o registro informará se a solicitação estava em *cache* ou não, ou se a requisição foi negada pelo servidor. A quinta coluna representa o tamanho (em *bytes*) da requisição. A sexta coluna representa o método de conexão do servidor (*GET* ou *POST*). Os três registros (ação, tamanho e método) não

apresentam utilidade para a abordagem proposta e por este motivo são descartados no momento de recuperação do *log*.

A informação mais importante para o ambiente proposto é a consulta. A consulta está localizada na sétima posição do registro. Para coletar esta informação é necessário verificar os detalhes de cada linha de registro, e entender o que separa a consulta que foi realizada, das outras informações do registro. Como visto na maioria dos registros, alguns parâmetros HTTP em uma consulta diferem das requisições normais. É crucial coletar apenas as consultas advindas de um motor de busca, como Google, e então, separá-las do resto do registro. O ambiente busca por parâmetros que marcam o início e o final das consultas.

O Quadro 5-1 apresenta a série de URLs identificadas advindas de conexões realizadas por meio do motor de busca Google, bem como os parâmetros de início e fim, respectivamente, utilizados em cada tipo de URL para identificação de consultas.

**Quadro 5-1. URLs originadas de pesquisas realizadas no motor de busca Google.**

<b>URL</b>	<b>Parâmetros</b>
<b>http://www.google.com.br/search?nord=1</b>	“&q=”, “&tch=”
<b>&amp;q=Leitura%20de%20arquivo%20XML&amp;tch=[...]</b>	
<b>http://www.google.com.br/search?nord=1[...]</b>	“&q=”, “&oq=”
<b>&amp;q=Inst%C3%A2ncia+de+classes+em+java&amp;oq=[...]</b>	
	“&spe”, “&pbx”

Após coletados os dados de data e hora, IP e consulta, toda a informação coletada é gravada em um banco de dados. O resultado que se tem é uma lista de endereços de IP, datas e consultas realizadas pelos colaboradores.

No entanto, mesmo que estas consultas representem a necessidade de um colaborador, nem todas as palavras digitadas no motor de busca influenciam no significado da consulta. Como o objetivo é prover informação relevante para auxiliar na criação do curso, a solução para extrair as palavras importantes foi utilizar algumas técnicas de mineração de textos como Tokenização, Filtragem e *Stemming*.

De forma a obter todas as palavras utilizadas em um texto, um processo de tokenização é necessário, no qual a consulta armazenada em banco de dados é dividida em uma lista de palavras. Para ilustrar o resultado de um processo de tokenização é utilizado o seguinte exemplo de consulta:

“Como instanciar classes em Java J2EE”;

Após o processo de tokenização, é realizada a divisão da consulta em uma lista de palavras. O resultado é apresentado abaixo:

“Como”, “instanciar”, “classes”, “em”, “Java”, “J2EE”;

Uma vez que todas as palavras estão divididas em uma lista, é possível executar o algoritmo de Filtragem. Esse algoritmo possui uma base de palavras, as quais não apresentam influência direta no significado de uma consulta. O algoritmo cruza a lista de palavras separadas pela tokenização com as palavras previamente inseridas na base de palavras do algoritmo. Ao encontrar alguma combinação de palavras, essa palavra é retirada da lista de palavras geradas pela consulta. O resultado do processo de filtragem é apresentado abaixo:

“instanciar”, “classes”, “Java”, “J2EE”;

Por fim, dentro das técnicas utilizadas em mineração de texto, é realizado o processo de *Stemming*. Este processo tenta construir as formas básicas de palavras, por exemplo, retirando o “s” dos substantivos que estão no plural, colocando verbos na sua forma infinitiva e retirando outros sufixos. O objetivo de executar o processo de *Stemming* é identificar a forma inicial de cada palavra, que é ponto de partida para variações desta palavra. Desta forma é possível encontrar as palavras que possuem o mesmo significado ou são iguais. Após o processo de *Stemming*, cada palavra é representada pela sua forma inicial. O resultado do processo de *Stemming* é representado por:

“instanc”, “classe”, “Java”, “J2EE”;

O próximo passo é agrupar (*clustering*) estes registros. O algoritmo de clusterização escolhido para agrupar estas consultas é o CLIQUE. A razão para utilizar o algoritmo CLIQUE é que no ambiente proposto, busca-se coesão entre os elementos do mesmo grupo, e este algoritmo de clusterização provê este resultado.

Entretanto, para realizar o agrupamento das palavras é necessária uma forma de comparar os grupos entre si. Como apresentado no Capítulo 3, a área de estudo de mineração de textos apresenta vários algoritmos e processos de comparação para definir medidas de similaridade, entre os quais se destacam como mais utilizados para similaridade de textos a Distância Euclidiana, a Similaridade de Cosine e o Coeficiente de Similaridade de Jaccard (HUANG, 2008). Todos estes processos buscam homogeneidade entre grupos de palavras.

O Coeficiente de Jaccard é baseado em similaridade de vetores e a cada ocorrência coincidente nos dois vetores V1 e V2, é incrementado o contador que irá gerar o grau de similaridade. O conteúdo dos vetores é separado em três conjuntos diferentes:

- A – que contém somente os valores coincidentes entre V1 e V2;
- B – que contém somente os valores não coincidentes de V1;
- C – que contém somente os valores não coincidentes de V2;

O coeficiente de Jaccard é calculado por:

$$\text{Jaccard} = A / (A+B+C) \quad (2)$$

O cálculo de Jaccard foi implementado no mecanismo na linguagem de programação C# (C-SHARP) e é calculado da seguinte forma:

1. São criadas duas listas que receberão os grupos que serão comparados.
2. As duas listas são transformadas em conjuntos codificados para comparação.

3. O método de intersecção (Intersect) encontra as diferenças entre o grupo 1 e o grupo 2. O método de união (Union) encontra as reincidências entre os grupos 1 e 2.
4. É aplicada a fórmula (2) para obter o coeficiente de Jaccard.

O coeficiente resultante será a variação entre 0 (não similar) e 1 (similar). Testes pilotos foram realizados com pesquisas simuladas em um motor de busca e foi constatado que o coeficiente que realiza os agrupamentos sem rejeitar temas que possuem apenas uma palavra diferente das demais, foi o valor 0.7. Sendo assim, este foi o limiar utilizado no mecanismo proposto para considerar pesquisas como sendo similares, conforme o exemplo:

“C# Estrutura de Controller Assíncrono” = “C# Controller Assíncrono”

O mecanismo realiza a pesquisa no banco de dados que retorna uma lista de grupos de palavras. O algoritmo de agrupamento ao receber a lista, executa os seguintes passos:

1. Cria o primeiro *cluster* e insere o primeiro grupo de palavras retornado do banco de dados.
2. Faz a comparação do próximo grupo retornado do banco com os grupos já inseridos em clusters. Essa comparação é realizada tendo como base o coeficiente de Jaccard (0.7).
3. Caso seja similar ao grupo do cluster comparado, ou seja, se o grupo de palavras é 70% similar aos grupos do cluster comparado, o grupo é inserido no mesmo cluster. Caso contrário o mecanismo cria um novo cluster e insere o novo grupo.
4. Enquanto houver grupos retornando do banco de dados, retorna ao passo 2.

O resultado final deste componente é uma lista de temas ordenados pelos temas mais pesquisados pelos colaboradores. Após a obtenção destes temas, são sugeridos ao usuário por meio do ambiente colaborativo semântico (MENOLLI, 2013), quais temas estão sendo mais pesquisados na internet. Após a sugestão do tema, fica a critério do especialista a escolha deste tema.

### 5.2.1. Arquitetura do Mecanismo Identificador de Temas

Esta seção apresenta a arquitetura (Figura 5-6) que organiza os componentes do mecanismo implementado para coleta de pesquisas realizadas em motores de busca e apresentação em forma de temas.

A arquitetura é subdividida em três camadas: aplicação, processamento, e *log de servidor* e banco de dados. O último bloco apresentado na arquitetura, chamado de Conexão com a Internet, trata-se da conexão feita pelos colaboradores e suas pesquisas, por este motivo não é tratado como uma camada do mecanismo.

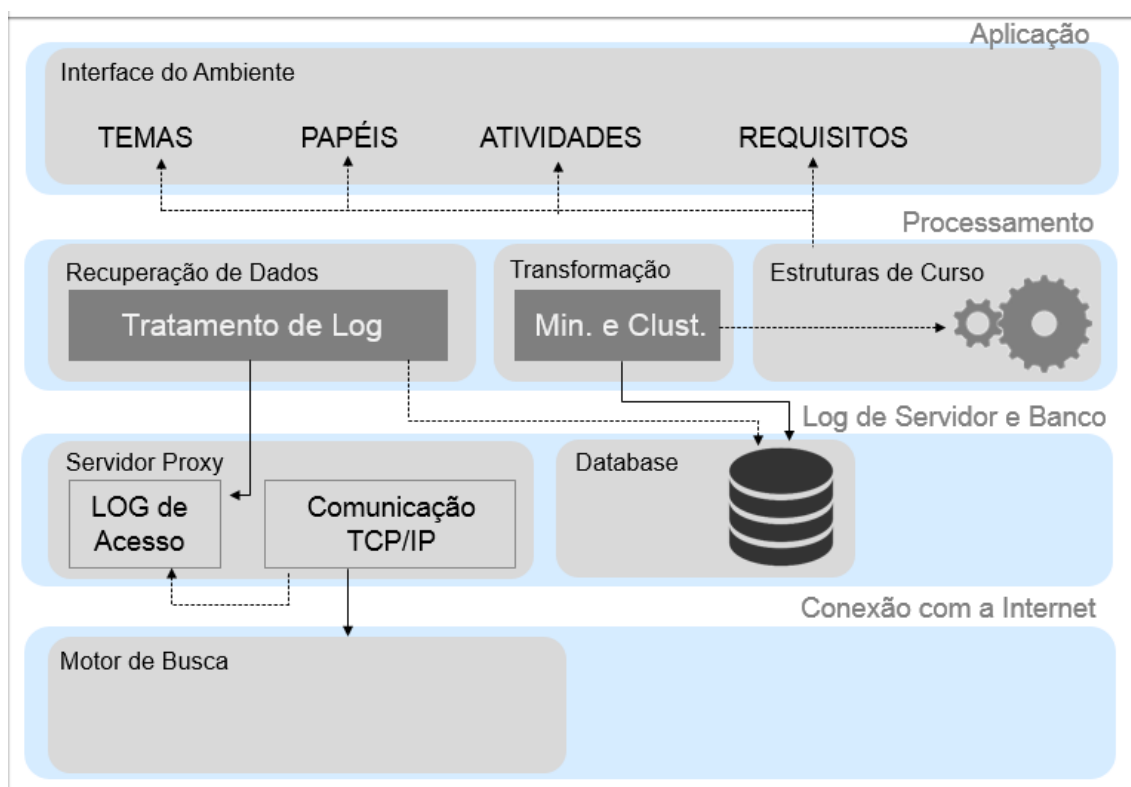


Figura 5-6. Arquitetura do Mecanismo Identificador de Temas.

A camada de aplicação é responsável pela interação do usuário e provê subsídio para a inclusão de conteúdo, como informações dos colaboradores e a criação da estrutura do curso. Esta camada é composta de quatro estruturas principais que podem ser alimentadas pela informação

coletada e tratada nas outras camadas. A estrutura definida pelo IMS Learning Design são: temas, papéis, conteúdos e hierarquia.

A camada de processamento central fornece uma combinação de técnicas de Recuperação de informação, mineração de texto e clustering. Além disso, a camada de processamento central realiza, como tarefa final, a transformação de estrutura de cursos, que coleta todos os dados agrupados e aplica outro algoritmo de mineração e clusterização, mas, identificando o tipo de estrutura de curso, por exemplo, “Exemplos”, “Conceitos”, “Exercícios”, etc. Também é aplicado um algoritmo para encontrar palavras que representem tecnologias que são as habilidades dos desenvolvedores, como por exemplo “Java”, “C#”, e “SQL”. Desta forma é possível utilizar estas palavras para criação de palavras chaves do curso, direcionar o curso para colaboradores específicos, etc.

Depois de processar estes dados, a camada central apresenta-os para a camada de aplicação, separando a informação coletada entre as estruturas do Learning Design, como “Tema”, “Papéis”, “Atividades”, e “Requisitos”.

A camada de banco de dados e de servidor de log é responsável por operar entre o usuário e a camada de conexão com a internet, coletando e armazenando as tentativas de conexão em um arquivo de log, e também armazenando a informação encontrada, depois de recuperá-la do servidor.

Assim, o objetivo desta arquitetura é a geração de estruturas de cursos, por meio de recuperação de informação compartilhada pelos membros da equipe em motores de busca da internet.

### **5.3 - Definidor de Curso**



O objetivo do Definidor de Curso (DC) é utilizar toda a informação obtida no mecanismo de identificação de temas e papéis para auxiliar o especialista a criar o curso e a partir deste curso, criar o esquema de aprendizagem.

Este módulo denominado Definidor de Curso foi dividido em três etapas. A primeira etapa é a de consulta de temas sugeridos, na qual o especialista poderá selecionar quais temas irão compor um curso. A segunda

etapa foi desenvolvida baseada em conceitos de modelagem de habilidades, que permite que o especialista defina quais as habilidades de um colaborador e qual o nível do colaborador em cada uma das habilidades. A terceira etapa de criação do curso e geração do esquema de aprendizagem foi desenvolvida baseada em conceitos educacionais que irão instruir o especialista no momento da criação do curso. Estes conceitos educacionais estão principalmente ligados teorias e modelos de design instrucional.

Para isto, foi implementado um ambiente para plataforma *web* que disponibiliza o módulo de geração de cursos. Na primeira etapa deste módulo - a etapa de consulta – um filtro de pesquisa é apresentado para o usuário. Neste filtro de pesquisa o usuário informa um período, que será utilizado pelo ambiente para retornar a lista de temas sugeridos com base nas pesquisas realizadas no período selecionado. A Figura 5-7 apresenta a etapa de consulta implementada para filtro dos temas sugeridos por período.

Relatório de Pesquisas

01/06/2015  03/06/2015 

Limpar **Pesquisar**

Tema	Pesquisas pelo Tema	
funções analíticas sql server	3	<input type="checkbox"/>
tabelas sys sql server	3	<input type="checkbox"/>
c# mvc controller actionresult jsonresult	3	<input type="checkbox"/>
linked server	2	<input type="checkbox"/>
c# sql data reader wcf service adapter	2	<input type="checkbox"/>
c# linha tendência	2	<input type="checkbox"/>
c# json request behavior allowget mvc3	2	<input type="checkbox"/>
region visual studio javascript	2	<input type="checkbox"/>
otimização consultas sql server	1	<input type="checkbox"/>
otimização arquitetura sql server	1	<input type="checkbox"/>
index online	1	<input type="checkbox"/>
mover tempdb novo diretório	1	<input type="checkbox"/>
gerenciando sql server	1	<input type="checkbox"/>

**Gerar Curso**

Figura 5-7. Tela para filtro de temas sugeridos por período.



Os objetos que estão sendo tratados neste ambiente são textos que representam a necessidade de aprendizagem do grupo. O trabalho do especialista é verificar e analisar a necessidade de aprendizagem do grupo e entender quais objetivos e métodos serão usados para auxiliar na melhora do conhecimento.

Na segunda etapa do módulo - a de identificação de habilidades de desenvolvedores de software – foram utilizados conceitos da área educacional que estão relacionados ao desenvolvimento de competências. O tema competência é constantemente discutido na literatura e empresas. Para Paquette (PAQUETTE, 2002), a Competência Profissional é a combinação de Conhecimentos, Habilidades e de Atitudes.

As competências em conhecimentos do profissional estão ligadas ao domínio de conhecimento, como por exemplo, conhecer um determinado assunto como a Engenharia de Software. Competências em habilidades são reconhecidas como as atividades que o profissional efetivamente consegue realizar, como por exemplo, desenvolver um sistema em uma determinada linguagem. Por fim, as competências ligadas às atitudes, estão ligadas ao comportamento do profissional em um ambiente empresarial.

O foco deste trabalho está em facilitar o desenvolvimento de habilidades de desenvolvedores de software. Em sua obra, Le Boterf (LE BOTERF, 1995), define que a competência do profissional, está ligada ao saber agir. Portanto, ao estruturar habilidades de desenvolvimento de software para criação de competências, o objetivo para o profissional dentro de uma organização se torna “Saber programar em uma linguagem”, ou “Saber gerenciar um projeto”, ou “Saber utilizar uma ferramenta de desenvolvimento”.

Desta forma, para poder identificar as habilidades dos desenvolvedores, e poder acompanhar o desenvolvimento destas habilidades, é proposta a utilização de uma identificação de habilidades específicas dos colaboradores da organização, bem como a classificação do nível deste conhecimento.

Basicamente, a tarefa de identificar o conhecimento organizacional consiste em identificar os conhecimentos relevantes, localizá-los e organizar essas informações de maneira que possam ser disponibilizadas para consulta. Esta identificação dos conhecimentos e classificação de níveis deve ser vista

como um guia, e deve apontar dentre os usuários, quais conhecimentos eles possuem, e em qual nível de entendimento.

A identificação das habilidades foi representada sobre a forma de uma tabela em um banco de dados, de modo que fique armazenado no servidor da organização, e assim, ao ser gerado um relatório dos logs do servidor Proxy, estes logs possam ser coletados com as informações sobre o conhecimento do usuário responsável. Foi implementado um módulo para manutenção das habilidades dos colaboradores, assim como o seu nível. Este módulo de manutenção das habilidades é utilizado quando o especialista necessita incluir um novo colaborador ou alterar os níveis de habilidades de um desenvolvedor.

Ao cadastrar um colaborador e selecionar uma habilidade, o especialista será solicitado a informar qual o nível do desenvolvedor. Os níveis incluídos no módulo são TRAINEE, JUNIOR, PLENO, SENIOR. Esta é uma representação padrão para nivelamento de cargos em organizações, por este motivo, esta classificação foi utilizada. Cabe ao especialista, seguindo um plano de cargos definido pelo setor de recursos humanos da organização definir qual o nível do desenvolvedor em determinada habilidade. A tela é apresentada na Figura 5-8.

Habilidades      Endereços IP

Habilidade: \* Seleccione... ▼

Nível: \* Seleccione... ▼

Incluir

Habilidade	Nível	
JAVA	PLENO	⊗
C#	SÊNIOR	⊗
SQLServer	JÚNIOR	⊗
JavaScript	PLENO	⊗
Objective-C	TRAINEE	⊗

Cancelar      Gravar

Figura 5-8. Tela de Cadastro de Habilidades de Desenvolvedores.

Para a terceira etapa do módulo Definidor de Curso - a de criação do esquema de aprendizagem - é necessário o envolvimento de um especialista, pois a tarefa de elaboração de um curso está ligada com a capacidade cognitiva do ser humano de interrelacionar conteúdos. No entanto, é necessário definir quem é o especialista que irá realizar a definição do curso.

Dentro de uma organização, normalmente não existe uma pessoa responsável pela criação de cursos e que é treinada para executar esta tarefa, geralmente o especialista é o colaborador com habilidades avançadas, ou um colaborador responsável pelos recursos humanos. Entretanto, estes colaboradores não possuem o conhecimento pedagógico educacional necessário para elaboração de cursos.

Desta forma, o ambiente gerador de cursos além de atuar como criador do Esquema de Aprendizagem, também deve auxiliar o especialista na elaboração do curso, e assim, oferecer unidades de aprendizagem mais efetivas para os colaboradores. Para isso, baseado nos conceitos apresentados no Capítulo 2, sobre modelos de design instrucional, foi proposta a combinação dos modelos SCM (Simplifying Conditions Method) (REIGELUTH, 1999) e SOI (Select, Organize, Integrate) (MAYER, 1996), para auxiliar o especialista durante o processo de criação de cursos. A combinação foi proposta nos três processos definidos pelo modelo SOI: seleção, organização e integração.

### Seleção

Os dois modelos, SOI e SCM, possuem a fase de seleção de temas. No modelo SCM, o processo de seleção sugere o levantamento de versões simples de execução de tarefas da habilidade que se deseja construir, progredindo para versões mais complexas. No modelo SOI, o processo de seleção, sugere a seleção apenas de informações relevantes que possam ser processadas pelo aprendiz.

A combinação dos modelos no ambiente proposto neste trabalho será relevante para auxiliar o especialista a selecionar temas que sejam relevantes para o aprendizado e que representem versões iniciais e simples das tarefas necessárias para o aprendizado da habilidade.

### Organização

O modelo SOI apresenta o processo de organização, enquanto que no modelo SCM, o processo de organização é implícito, pois, o modelo sugere que os conteúdos devem ser organizados de versões simples da tarefa, evoluindo para versões complexas. No ambiente proposto, a organização se dá pela apresentação de sugestões que levem o especialista a organizar o conteúdo de forma coerente.

### Integração

O último processo, o de integração, não é previsto pelo modelo SCM, sendo utilizado apenas no modelo SOI, que sugere que o aprendizado deve estar conectado com o conhecimento já adquirido pelo aprendiz. No modelo proposto a integração se dá pela seleção dos colaboradores. O especialista receberá sugestões dos colaboradores que tenham suas habilidades relacionadas com os temas selecionados de forma a dar continuidade no aprendizado do colaborador.

Estes processos (Seleção, Organização e Integração) foram implementados no ambiente gerador de Esquema de Aprendizagem. Com o objetivo de auxiliar o especialista na tarefa de criação do esquema, foram propostas quatro etapas para geração do curso. Desta forma, na maioria das etapas foi possível implementar um processo do modelo SOI. As etapas, bem como o que foi implementado em cada uma delas é apresentado na Seção 5.3.1.

#### **5.3.1. Criação do Esquema de Aprendizagem**

Os cursos são elaborados de maneira sequencial, combinando os modelos SOI e SCM, em quatro passos principais.

1. Coleta de informações gerais: neste passo o especialista será solicitado a informar o título do curso, subtítulo, objetivo, requisitos e palavras chaves que possam auxiliar na localização do curso. A tela correspondente a este passo é apresentado na Figura 5-9.

Gerador de Curso

1. **Informações Básicas**    2. Temas    3. Organização    4. Papéis    Anterior    Próximo

Título:

Subtítulo:

Objetivo:

Key Words:

Requisitos:

Descrição:

Key Words:

**Figura 5-9. Etapas para Criação do Curso. Passo 1 - Informações Básicas.**

As informações coletadas nesta etapa representam basicamente o cabeçalho do Esquema de Aprendizagem, representado pelas estruturas *Learning Design* e *Prerequisites*.

Informações como pré-requisito e objetivos podem ser apresentadas automaticamente no ambiente, utilizando as informações dos temas selecionados. Por exemplo, ao selecionar os temas "C# desenvolvimento MVC" e "C# classes assíncronas", o ambiente irá assumir que os palavras chave para este curso são "C#" e "MVC", por serem palavras que estão ligadas às habilidades dos desenvolvedores.

2. Seleção e criação de novos temas: mesmo que o especialista possa selecionar no início da criação do curso alguns dos temas que foram sugeridos pelo mecanismo, como apresentado na Figura 5-7, neste passo, o especialista é convidado a escolher novos temas que estão relacionados aos temas selecionados previamente, e também a criar novos temas. Também neste passo, é iniciado o processo de auxílio na elaboração do curso com base na combinação dos modelos SOI (MAYER, 1996) e SCM (REIGELUTH, 1999). Por se tratar da etapa de seleção de temas, o ambiente irá apresentar dicas de como o especialista pode selecionar os temas, para que haja maior efetividade na aprendizagem do curso. A etapa 2, de seleção e criação de temas é apresentada na Figura 5-10.

Gerador de Curso

1. Informações Básicas    **2. Temas**    3. Organização    4. Papéis    [Anterior](#)    [Próximo](#)

controller actionresult jsonresult 4

linha tendência criando constantes

request behavior allowget

Novo Tema

Temas Relacionados

- c# mvc controller actionresult jsonresult 4
- c# linha tendência criando constantes
- c# json request behavior allowget mvc3
- c# sql data reader wcf service adapter
- c# mvc partial view model
- c# visual studio multiplos projectos projetos
- c# command parameters add remove
- ajax call function controller c# mvc parameters
- post json controller mvc get
- web config c# key string
- c# mvc diferença motores razor aspx
- c# mcv 4 razor
- c# mvc razor master content desempenho

**SELEÇÃO**

DICAS DE SELEÇÃO DE TEMAS RELEVANTES

**Figura 5-10. Etapas de Criação do Curso. Passo 2 - Temas**

As informações inseridas nesta etapa serão utilizadas pelo próximo passo, que é a organização e hierarquia de temas.

3. Organização dos temas: nesta etapa o especialista será solicitado a organizar os temas selecionados na etapa 2. Seguindo também a combinação dos modelos SOI (MAYER, 1996) e SCM (REIGELUTH, 1999), o ambiente apresenta dicas que irão auxiliar o especialista na organização dos temas. A tela desta etapa é apresentada na Figura 5-11.

Gerador de Curso

1. Informações Básicas    2. Temas    3. Organização    4. Papéis

Anterior    Próximo

request behavior allowget

- controller actionresult jsonresult 4

linha tendência criando constantes

ORGANIZAÇÃO

DICAS DE ORGANIZAÇÃO DOS TEMAS SELECIONADOS

**Figura 5-11. Etapas de Criação do Curso. Passo 3 - Organização**

Os temas do Esquema de Aprendizagem serão preenchidas automaticamente. Estas informações no Esquema de Aprendizagem são responsáveis pela formação da hierarquia do curso, em temas e subtemas. Um exemplo de hierarquia de temas é apresentado na Figura 5-12, no qual os temas estão organizados em temas e subtemas.

request behavior allowget

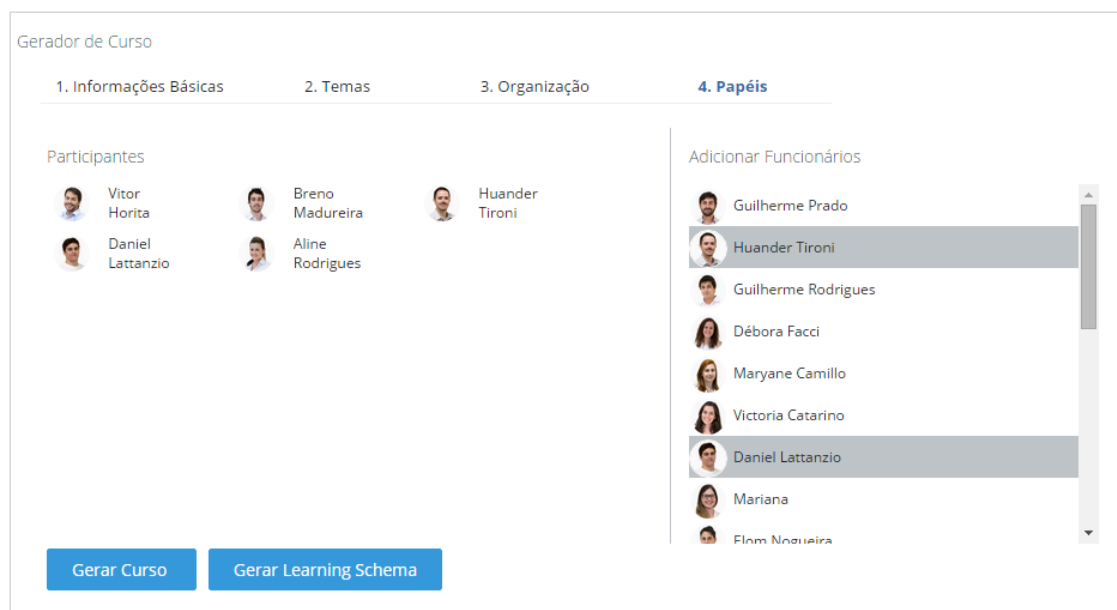
- controller actionresult jsonresult 4

linha tendência criando constantes

**Figura 5-12. Estrutura dos temas.**

4. Seleção de papéis: nesta etapa o especialista é solicitado a informar quais colaboradores devem participar do curso. Esta etapa também foi adequada de acordo com o processo SOI (MAYER, 1996), no que diz respeito à integração das informações. O ambiente contribuirá com a criação do curso, apontando possíveis participantes para o curso, que serão escolhidos de acordo com os requisitos do curso e as habilidades dos desenvolvedores. Assim, o especialista poderá integrar os habilidades que deverão ser construídas a partir do curso criado, com as habilidades existentes

dos colaboradores. A etapa de seleção de papéis é apresentada na Figura 5-13.



**Figura 5-13. Etapas de Criação do Curso. Passo 4 - Papéis**

Nesta etapa, as informações preenchidas automaticamente no Esquema de Aprendizagem são *Learner e Staff*. Na estrutura de *learner* são adicionados os colaboradores que serão aprendizes, e na estrutura de *staff* os colaboradores que são considerados aptos a orientar o curso.

A tarefa final do Definidor de Cursos é gerar o Esquema de Aprendizagem em um arquivo de formato XML, que possa ser utilizado pelo especialista para gerar o curso com seus conteúdos em um ambiente como o proposto em Menolli, Malucelli e Reinehr (MENOLLI; MALUCELLI; REINEHR, 2011).

Ao selecionar a opção de “Gerar Learning Schema”, como apresentado na Figura 5-13, o ambiente salvará as informações de criação de curso em um banco de dados, e irá gerar o Esquema de Aprendizagem. Este esquema será montado no arquivo com extensão .xml, que será baixado pelo especialista.



## **5.4 - Considerações Finais**

Este capítulo apresentou as principais abordagens propostas para este trabalho e a suas respectivas implementações. Para a primeira abordagem, o ITP (Identificador de Temas e Papéis), também foi apresentada uma arquitetura que descreve o funcionamento do mecanismo. Para a abordagem de Definição de Cursos, foi apresentado como um processo passo a passo é utilizado para coletar informação necessária para criação do Esquema de Aprendizagem, e também prover um modelo baseado em conceitos educacionais para criação de cursos. Por fim, foi apresentado o objetivo final do ambiente Definidor de Cursos, que é a criação do Esquema de Aprendizagem. Nos próximos Capítulos, 6 e 7, foram apresentados experimentos que foram realizados para avaliar a proposta.

## Capítulo 6 - **EXPERIMENTO PARA AVALIAR A IDENTIFICAÇÃO DE TEMAS**

Este capítulo tem como propósito principal apresentar o experimento conduzido com uma equipe de desenvolvimento de software, com o objetivo de testar e avaliar o mecanismo desenvolvido para captar as consultas realizadas em motores de busca. Para atingir esta meta, primeiramente são apresentados o objetivo geral e os objetivos específicos, tipo e planejamento do experimento de identificação de temas relevantes. Também foi apresentado o método de coleta dos dados, e por fim seus resultados e discussão.

### **6.1 - Introdução**

Uma arquitetura que suporta este mecanismo foi proposta, e a partir desta arquitetura, um esquema, que representa a abordagem para criação do Esquema de Aprendizagem foi exposto. No entanto, antes de realizar a implementação completa do mecanismo, a fim de confirmar a viabilidade da proposta, foi realizado um experimento com uma equipe de desenvolvimento de software, para avaliar a viabilidade de utilização da abordagem de coleta de temas e papéis de desenvolvedores. Assim, foi definido um motor de busca para coleta de dados, e a partir disto, foi realizado o experimento para avaliação do identificador de temas e papéis.

### **6.2 - O experimento**

Este experimento foi proposto para avaliar como as consultas realizadas pelos colaboradores da equipe de desenvolvimento podem auxiliar na descoberta e sugestão de temas relevantes para a aprendizagem dos membros da equipe.

Como a utilização de um motor de busca na abordagem proposta, é um passo necessário para coletar pesquisas, e o motor de busca Google possibilita a coleta por meio do registro de *Log*, este motor de busca foi utilizado. Dessa

maneira, foi utilizada uma conexão em rede com internet, de forma que os participantes do experimento tivessem acesso ao motor de busca para registrar as pesquisas.

O Google foi selecionado como fonte de coleta de pesquisas neste trabalho, por ser o motor de busca mais utilizado na maioria dos países. De acordo com o NetMarketShare (NETMARKETSHARE, 2015), 65,7% de usuários de motores de busca, utilizam o Google. Outra vantagem é que de forma a proteger as pesquisas de seus clientes, o Google utiliza como padrão, o protocolo HTTPS para encriptação das consultas, no entanto, é disponibilizado outro link de acesso – o [www.google.com.br/?nord=1](http://www.google.com.br/?nord=1) – que não realiza o direcionamento da consulta do usuário para o protocolo HTTPS, tornando possível o registro dos *logs* de pesquisa em um servidor *proxy*.

Uma rede interna foi montada de forma que as conexões realizadas para consultas no motor de busca pelos usuários tramitassem primeiramente pelo servidor *Proxy*. A intenção foi analisar o uso do mecanismo em um ambiente controlado e validar o funcionamento da abordagem implementada.

### **6.3 - O Método Experimental**

O presente experimento é caracterizado como um estudo experimental, que visa coletar dados em um ambiente controlado, para confirmar ou negar uma hipótese. O método experimental considera a proposta e avaliação de um modelo com os estudos experimentais (BASILI, 1996). O paradigma experimental envolve um projeto, observação, coleta de dados e validação sobre o processo ou produto que está sendo estudado.

O método experimental sugere o modelo, desenvolve o método qualitativo e/ou quantitativo, aplica um experimento, mede e analisa, avalia o modelo e repete o processo. Isto é uma abordagem orientada à melhoria evolucionária. O processo se inicia com o levantamento de um modelo novo, não necessariamente baseado em um modelo já existente e tenta estudar o efeito do processo ou produto sugerido pelo modelo novo (TRAVASSOS; GUROV; AMARAL, 2002). É importante ressaltar que experimentos não provam nada, nenhum experimento oferece prova com certeza absoluta. Os

experimentos verificam a previsão teórica da realidade (BASILI, 1996). A maior vantagem do experimento é o controle total sobre o processo e as variáveis, além da possibilidade de repeti-lo.

Portanto, pode-se classificar este experimento de acordo com Basili (BASILI, 1996) como:

- Estudo descritivo, pois pode haver padrões nos dados, mas a relação entre as variáveis não foi analisada.
- In vitro, pois foi executado em uma empresa de desenvolvimento de software, no entanto, sob condições controladas.
- Aplicado em especialistas, pois os praticantes possuem experiência no domínio de estudo como apresentado nas seções seguintes nos requisitos do experimento.
- Estudo Observacional, uma vez que não há nenhum tratamento ou variáveis controladas.

#### **6.4 - Definição dos Objetivos**

O objetivo geral deste experimento é analisar se os temas sugeridos pelo mecanismo, com base em pesquisas realizadas pela equipe de desenvolvimento de software, representam a necessidade de aprendizado geral do grupo de participantes. Para atingir este objetivo foram delineados alguns objetivos específicos que deveriam ser realizados dentro do experimento:

- Apresentar o mecanismo identificador de temas e papéis a uma equipe de desenvolvimento de software;
- Caracterizar a validade da aplicação da abordagem em um ambiente empresarial na visão dos colaboradores;
- Analisar os temas sugeridos.

#### **6.5 - Planejamento**

O planejamento do experimento foi dividido em cinco etapas. Cada etapa é descrita nas seções seguintes.

### **6.5.1. Definição das hipóteses**

Para orientar a pesquisa, três hipóteses foram estabelecidas com base no objetivo principal deste experimento, que são descritas abaixo:

Hipótese (H1): os temas sugeridos pelo mecanismo representam a necessidade de aprendizagem da maioria dos membros da equipe de desenvolvimento.

Hipótese (H2): a clusterização realizada pelo mecanismo das pesquisas efetuadas pelos desenvolvedores contribui para que os temas sugeridos não se repitam e sejam considerados relevantes.

Hipótese (H3): a abordagem proposta para coleta de informações de consulta realizadas em motores de busca, não é rejeitada pelo grupo de desenvolvedores.

### **6.5.2. Seleção dos participantes**

Primeiramente é necessário delimitar as características mínimas para a escolha dos participantes do experimento. Como o estudo se refere especificamente a empresas desenvolvedoras de software, foi proposta a utilização de funcionários de uma empresa de desenvolvimento de software do ramo de rastreamento veicular.

Foram selecionados quatro colaboradores para serem participantes do experimento, dentre os quais todos estão de alguma forma ligados ao desenvolvimento de software. Por se tratar de uma equipe de desenvolvimento relativamente pequena, cada participante desempenha um papel diferente dentro do grupo. Todos eram do sexo masculino e tinham entre 20 e 25 anos de idade e no mínimo 2 anos de experiência na área de desenvolvimento de software, seja com administração de banco de dados ou programação.

### 6.5.3. Descrição do Experimento

Neste experimento, o principal objetivo é avaliar se a abordagem proposta, de coletar os dados de consultas realizadas em motores de busca e depois apresentá-los em forma de temas, pode auxiliar o especialista na criação de cursos, por meio de buscas realizadas pelos participantes.

De forma a gerenciar a informação sobre o participante responsável pela consulta, foi implementada uma tela que permite ao especialista Criar, Visualizar, Atualizar e Deletar colaboradores, bem como seus conhecimentos específicos, níveis, e o endereço IP utilizado por este participante na rede montada.

De modo a simular um ambiente organizacional, uma rede local foi construída. Também foram criados quatro colaboradores fictícios, sendo um para cada participante do experimento. Cada colaborador foi mapeado com um endereço IP da rede local, de forma que o mecanismo possa encontrar o participante responsável pela pesquisa.

Depois de criados os quatro colaboradores, foi necessário coletar as suas respectivas pesquisas realizadas no motor de busca Google. Os participantes foram solicitados a realizar 50 consultas relacionadas a desenvolvimento de software em um motor de busca. Os critérios para realizar as consultas eram que o assunto deveria estar relacionado com os projetos da empresa e as habilidades do desenvolvedor.

Depois disso, as consultas dos pesquisadores foram recuperadas do *log* do servidor *Proxy* e armazenadas em um banco de dados. A partir disso, o componente de Transformação de Dados, com base nas consultas armazenadas, criou uma lista com listas de grupos de palavras.

Após o procedimento de agrupamento, as consultas foram apresentadas como temas pelo mecanismo, com a quantidade de vezes que a consulta ao tema foi realizada pelos colaboradores.

Os resultados dos temas sugeridos foram coletados para análise dos dados. Além disso, cada participante ao fim desta tarefa, respondeu um questionário para expressar sua opinião sobre a abordagem.

#### 6.5.4. Descrição da Análise

As áreas abrangentes do desenvolvimento de software compreendidas foram a de programação de interfaces e regras de negócio, análise e programação de banco de dados, e administração de banco de dados.

Foi solicitado a cada colaborador, que informasse no mínimo 50 temas de pesquisas que eles poderiam realizar em um motor de busca. O critério para realizar as buscas, foi que o assunto estivesse relacionado com os projetos da companhia e competências do colaborador.

Para coletar a informação sobre o usuário responsável pela busca, estes funcionários foram cadastrados no ambiente com suas habilidades específicas e os endereços de IP que cada funcionário utiliza para navegação. Neste experimento os usuários foram caracterizados como Colaborador e o seu número de cadastro, por exemplo: Colaborador 1 (C.1). Após criado os 4 colaboradores, foi necessário realizar as buscas informadas pelos colaboradores no motor de busca escolhido para o experimento, o Google.

Uma rede empresarial foi simulada de forma que as pesquisas informadas por cada colaborador fossem realizadas a partir de um computador diferente. Cada computador teve sua conexão apontada para um servidor *proxy* responsável em coletar as informações de cada requisição e registrar em um arquivo de *log*.

Posteriormente, o mecanismo foi utilizado para recuperar os registros do *log* referentes às pesquisas realizadas. Estes registros de pesquisas recuperados foram armazenados em um banco de dados. Uma vez que os registros eram recuperados do arquivo de *log* e armazenados, os registros do *log* eram apagados para evitar que o servidor fosse sobrecarregado sempre que tentasse recuperar novos registros. Em seguida, o componente de Transformação de Dados, apresentado no Capítulo 5, foi utilizado para criar, a partir das consultas armazenadas no banco de dados, uma lista de grupos de palavras. Ao todo foram formados 150 grupos de palavras, nos quais, cada grupo corresponde a uma pesquisa realizada pelos participantes. Cada grupo conteve em média quatro palavras.

O componente de Transformação de Dados, que também tem como objetivo utilizar algoritmos de mineração de textos, realizou o processo de

mineração que é composto por quatro etapas: tokenização, filtragem, stemming e filtragem de tecnologias. Cada etapa foi realizada nos grupos individualmente. A primeira etapa de tokenização, como apresentado no Capítulo 3, divide o grupo de palavras em palavras isoladas. Ao final do processo de tokenização, os grupos formaram 652 palavras isoladas. Este processo é a base para realizar a filtragem.

Na segunda etapa de filtragem, o componente de Transformação de Dados comparou cada palavra isolada dos 150 grupos formados com uma lista de palavras denominadas *stop words*, e retirou dos grupos as palavras que constavam nessa lista. As *stop words* definidas para o algoritmo foram apresentadas no Capítulo 5. Ao todo, das 652 palavras isoladas, 114 foram retiradas na filtragem, restando um total de 538 palavras.

A terceira etapa do processo de mineração de textos, pelo qual as palavras restantes dos grupos foram submetidas foi a etapa de *stemming*. Como este processo não visa retirar palavras dos grupos, a quantidade de palavras permaneceu a mesma. No entanto, é importante apresentar o resultado de algumas palavras resultantes do processo. Na Tabela 6-1 são apresentadas dez palavras que foram informadas pelos participantes e passaram pelo processo de *stemming*.

Tabela 6-1. Exemplos de palavras antes e após o processo de Stemming.

Palavra original	Resultado do Stemming
<b>Funções</b>	Fun
<b>Analíticas</b>	Analitic
<b>Otimização</b>	Otimiz
<b>Diretório</b>	Diretor
<b>Gerenciando</b>	Gerenc
<b>Estrutura</b>	Estrut
<b>Paginação</b>	Pagin
<b>Manutenção</b>	Manuten
<b>Particionamento</b>	Partic
<b>Utilização</b>	Utiliz

Por fim, a quarta e última etapa da mineração de texto realizada foi retirar dos grupos as palavras que de representassem estruturas de cursos,



como os exemplos apresentados no Capítulo 5. Estas palavras foram retiradas dos grupos, no entanto foram incluídas em listas nas quais permaneciam relacionados aos seus grupos de origem para posteriormente serem reutilizadas.

A análise deste experimento foi realizada utilizando análise qualitativa. O estudo qualitativo está relacionado com a pesquisa sobre os objetos quando os resultados são apresentados em termos naturais (SEAMAN, 1999).

A análise foi realizada por meio da análise textual discursiva (MORAES; GALIAZZI, 2006). Esta análise é recomendada na pesquisa qualitativa, e compartilha muitas das premissas evidentes em outras metodologias que também pertencem ao campo da análise textual, como a análise de conteúdo (BARDIN, 1977) e análise do discurso (PÊCHEUX, 1995).

Para realizar a análise, a fim de responder às hipóteses propostas na seção 6.5.1, um questionário com questões fechadas foi aplicado aos colaboradores da equipe de desenvolvimento, de forma a avaliar a aplicação da abordagem de coleta de temas e papéis de desenvolvedores em um ambiente empresarial. O questionário é apresentado no Quadro 6-1.

**Quadro 6-1. Questionário aplicado no experimento para avaliação da abordagem.**

<b>Questão</b>	<b>Descrição</b>
<b>Q1</b>	Você utiliza sites de busca como Google, Yahoo e Bing para encontrar resoluções para as dificuldades do dia a dia no âmbito de desenvolvimento de sistemas?
<b>Q2</b>	A empresa contribui com a sua pesquisa diária deixando liberado na rede os acessos a fóruns, blogs ou wikis?
<b>Q3</b>	Você no dia a dia procura respostas para dúvidas pontuais (pequenas dúvidas que impedem o andamento da tarefa) primeiramente com os colegas de trabalho antes de fazer a pesquisa na Internet?
<b>Q4</b>	A seguinte busca foi realizada por um ou mais respondentes: "C# MVC Controller Action Result Json". Os temas envolvidos nesta busca tem ligação com as tecnologias do seu dia a dia?
<b>Q5</b>	Você considera que é importante a empresa ter conhecimento sobre quais as dúvidas o funcionário tem durante a realização de uma tarefa?
<b>Q6</b>	Você evitaria fazer as buscas de dúvidas na internet se soubesse que elas estão sendo coletadas para melhor entendimento das dificuldades e criação de cursos?
<b>Q7</b>	Você acredita que executaria de maneira mais eficaz seu trabalho se houvessem cursos com temas voltados para os projetos internos?
<b>Q8</b>	Neste experimento cerca de 200 pesquisas diferentes foram registradas. Dentre

---

as pesquisas o seguinte resultado foi encontrado: "Index Online SQL Server".  
Você considera este assunto importante no seu dia a dia?

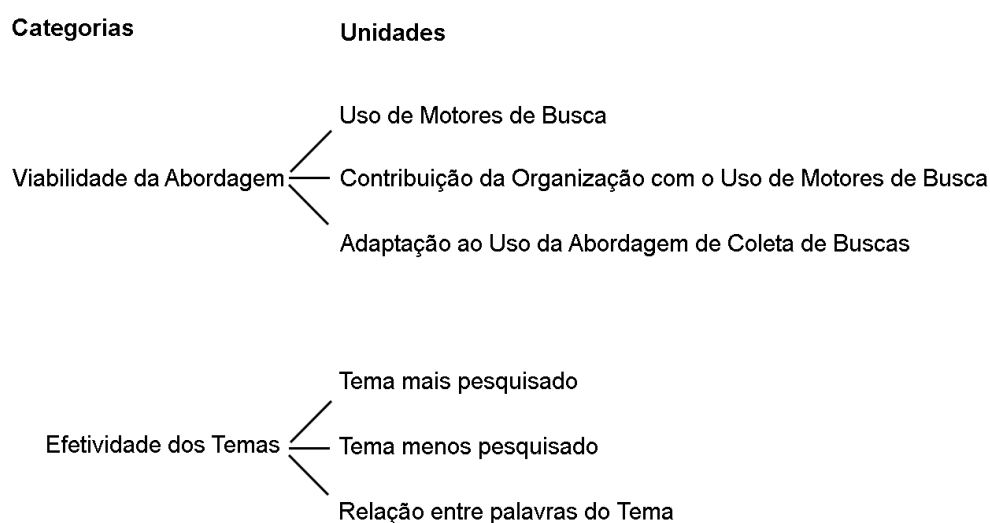
**Q9** Dentre os temas sugeridos para criação de cursos, foi sugerido "C# Json Request Behaviour AllowGet MVC3". Você concorda que há relação entre todas as palavras do tema?

**Q10** Você se considera apto para lecionar um curso dentro da empresa com o seguinte tema: "C# SQL DataReader"?

---

Assim, o primeiro passo da análise é definir as categorias, que são as unidades de âmbito geral, estabelecendo limites para a análise (MORAES; GALIAZZI, 2006). Cada categoria geralmente contém várias unidades de análise, que é o elemento unitário de conteúdo a ser submetido à classificação (MORAES; GALIAZZI, 2006). Portanto, as unidades de análise são as entidades que são analisadas no estudo.

Assim, a estrutura geral da análise textual discursiva é mostrada na Figura 6-1, que apresenta seis unidades de análise organizadas em duas categorias.



**Figura 6-1. Estrutura Geral da Análise Discursiva Textual.**

Após a coleta das pesquisas realizadas pelos desenvolvedores e suas respectivas competências, os questionários respondidos pelos participantes foram analisados. Para facilitar a compreensão da análise textual, foram utilizados os seguintes identificadores simbólicos:

- C1, C2, C3, C4: identifica o colaborador participante do experimento;

- Q1, Q2, Q3, ..., Q10: identifica as perguntas do questionário aplicado após a conclusão da coleta de pesquisas no motor de busca.

### **6.5.5. Execução**

Como mencionado na Seção 6.5.3, uma rede local foi construída. Cada desenvolvedor participante do experimento foi mapeado em um banco de dados, juntamente com suas competências. Os quatro colaboradores foram denominados C.1, C.2, C.3 e C.4. Além das competências, cada colaborador foi mapeado com um endereço IP da rede local, de forma que o mecanismo possa encontrar o participante responsável pela pesquisa.

Os participantes foram solicitados a realizar 50 consultas relacionadas a desenvolvimento de software no motor de busca Google. Os critérios para realizar as consultas, deveriam estar relacionados com os projetos da empresa e as habilidades do desenvolvedor.

O ambiente realizou a coleta das pesquisas e agrupamento dos temas e apresentou na tela de filtro de temas apresentado na Figura 5-7, com a quantidade de vezes que a consulta ao tema foi realizado pelos colaboradores e a possibilidade de seleção de temas para criação de cursos.

O experimento para avaliação da etapa de criação de cursos foi abordado no Capítulo 7.

Nas seções seguintes são apresentados os resultados da análise textual para os temas apresentados pelo ambiente e para cada questão apresentada no Quadro 6-1.

## **6.6 - Resultados**

Nesta seção os principais resultados após a análise das técnicas utilizadas na abordagem, bem como os temas coletados são apresentados. Também serão apresentados os resultados referentes a cada categoria e unidades de análise. Para cada unidade de análise, as perguntas respondidas pelos participantes foram analisadas, explorando as principais informações.

### 6.6.1. Mecanismo de Sugestão de Temas

A primeira tecnologia analisada foi a utilização do motor de busca Google, posteriormente foi avaliado o método de recuperação de informações do *log* e as técnicas utilizadas para realizar a mineração de textos. Por fim, foi analisada a técnica de clusterização de textos. As análises são apresentadas a seguir:

#### 1. Motor de busca Google

A utilização do motor de busca Google para o experimento se mostrou efetivo, pois, por ser uma ferramenta que os participantes já estavam familiarizados, não se tornou uma barreira para as pesquisas.

Como discutido no Capítulo 5 na Seção 5.2, poderiam ser utilizadas outras formas de coleta de necessidades de aprendizagem de desenvolvedores, como por exemplo, a criação de um *plug-in* para navegadores no qual o desenvolvedor informasse diretamente a sua necessidade de aprendizagem, ou a criação de um ambiente para realizar pesquisas na internet, entre outras. No entanto, estas abordagens, quando aplicadas em um ambiente corporativo real, demandariam atenção dos desenvolvedores em informar suas dúvidas em uma ferramenta diferente de um motor de busca. O motor de busca é o principal meio pelo qual o desenvolvedor recorre às suas dúvidas (PRASATH et al., 2014). Portanto a utilização de motores de busca para a coleta destas pesquisas foi considerada uma abordagem efetiva para este trabalho.

Entretanto, o motor de busca Google, de forma a garantir a segurança da informação das pesquisas realizadas pelos usuários, sempre redireciona as pesquisas para um protocolo HTTPS, que faz a criptografia das informações digitadas. Para que o servidor *proxy* implementado pudesse coletar as pesquisas realizadas, foi necessário solicitar que os participantes realizassem as buscas utilizando uma URL específica do Google (<http://www.google.com.br/?nord=1>).

Se aplicado num ambiente empresarial real, no qual não seja possível controlar as ações diárias, seria necessária uma adequação de processos

organizacionais para utilização desta URL específica para possibilitar a coleta das necessidades de melhoria de competências.

## 2. Mineração de textos

Em relação às técnicas de mineração de textos aplicadas, inicialmente para as aplicadas com objetivo de processar o texto coletado das buscas, foi possível concluir que o primeiro processo, o de tokenização, é essencial para a que as próximas fases sejam realizadas de forma a conseguir resultados relevantes.

Para os processos seguintes, de filtragem e *stemming*, foi possível notar pontos nos quais pudessem ocorrer melhorias. Os processos normais de filtragem e *stemming* de textos encontrados na literatura são referentes a conteúdos nos quais são mais descritivos e dissertativos. O foco da mineração de textos deste trabalho é referente a pesquisas realizadas em motores de busca, nos quais as palavras digitadas pelos desenvolvedores representam tecnologias e termos técnicos em inglês e português. Por este motivo, foi necessário complementar a implementação do algoritmo de filtragem e *stemming* com regras de processamento de palavras específicas que representassem tecnologias.

A melhoria executada no mecanismo nesta etapa foi a coleta de palavras específicas de tecnologias como, por exemplo, C#, Java, MVC, e Javascript, que foi definido como Coletor de Requisitos Técnicos (CRT) e apresentado no Capítulo 5. O objetivo do CRT é descobrir quais são as tecnologias pesquisadas pelos desenvolvedores, e utilizá-las para definição de requisitos na próxima etapa, que é a criação de cursos.

As consultas digitadas pelos participantes foram objetivas e continham poucas palavras que passassem pelo tratamento de filtragem como a retirada de *stop words*. Pelo mesmo motivo, algumas palavras relacionadas a tecnologias, termos técnicos de origens do inglês, os quais não foram previstos no tratamento de palavras específicas, foram vistos como melhorias a serem implementadas no algoritmo. As palavras identificadas neste experimento foram incluídas no algoritmo do CRT, no entanto, é evidenciada neste trabalho como uma limitação, a necessidade de incluir manualmente palavras novas.

### 3. Clusterização das pesquisas

O agrupamento das pesquisas, com objetivo de encontrar os temas mais relevantes, é uma tarefa essencial. Desde o início da escolha do algoritmo de clusterização e implementação, precauções foram tomadas para evitar que temas que pudessem compor o mesmo curso, fossem tratados como temas diferentes.

Nas 200 pesquisas realizadas pelos participantes do experimento foi verificada a média de palavras digitadas no motor de busca. O resultado mostrou que a média de palavras utilizadas eram 4 por pesquisa. Pesquisas nas quais os temas eram os mesmos, poderiam estar representadas por 5 ou 3 palavras, dependendo do detalhamento informado pelo desenvolvedor, mesmo após o processamento do texto. Por isso, a solução encontrada foi na escolha do algoritmo de similaridade utilizado para comparar as consultas. O coeficiente de indexação de Jaccard, como apresentado no Capítulo 3, permite variações de 0 à 1, sendo consideradas dissimilares quando o valor é mais próximo de 0, e similares quando o valor é mais próximo de 1.

Sendo assim, a comparação de similaridade entre uma frase que possui 5 palavras e outra com 3, considerando o máximo possível de palavras iguais nas duas frases que é 3 palavras, resultaria em um coeficiente de igualdade de 0.6 (60%). Ou seja, considerando os exemplos de consultas a seguir que foram coletados das pesquisas realizadas no experimento:

- “C# JsonResult ActionResult”
- “MVC3 Implementar JsonResult ActionResult C#”

Para o exemplo apresentado, mesmo com assuntos pertencentes ao mesmo tema, eles foram considerados com apenas 60% de similaridade. Por esse motivo, após vários testes, com pesquisas diferentes, o coeficiente definido previamente no desenvolvimento do ambiente e apresentado no Capítulo 5 como 0.7 (70%) foi reajustado para 0.6 (60%), para que o ambiente possa realizar melhor a clusterização dos temas.

Um possível trabalho futuro já identificado para esta abordagem é a utilização de um limiar de similaridade adaptativo, que seja capaz de, de acordo com a quantidade de palavras, encontrar o melhor valor de coeficiente a ser utilizado para comparação entre as pesquisas.

## 6.6.2. Unidades de Análise

A primeira categoria analisada foi Viabilidade da Abordagem. Os resultados de cada unidade de análise desta categoria encontram-se descritos a seguir.

### 1. Uso de Motores de Busca

O objetivo desta unidade é responder às perguntas: “Você utiliza sites de busca como Google, Yahoo e Bing para encontrar resoluções para as dificuldades do dia a dia no âmbito de desenvolvimento de sistemas?” e “Você no dia a dia procura respostas para dúvidas pontuais (pequenas dúvidas que impedem o andamento da tarefa) primeiramente com os colegas de trabalho antes de fazer a pesquisa na Internet?”. A análise desta unidade foi feita com base nas respostas dos participantes.

Considerando as respostas dos participantes, 100% utilizam motores de busca frequentemente e priorizam a utilização de motores de busca quando as dúvidas são referentes a tecnologias de desenvolvimento de *software*.

### 2. Contribuição da Organização com o Uso de Motores de Busca

O objetivo desta unidade é responder à pergunta: “A empresa contribui com a sua pesquisa diária deixando liberado na rede os acessos a fóruns, blogs ou *wikis*?”. Para acessos à internet cujo objetivo é para resolução de problemas relacionados ao desenvolvimento de software, todos os participantes responderam que a empresa contribui para a pesquisa liberando o acesso a fóruns, *blogs* ou *wikis*.

### 3. Adaptação ao Uso da Abordagem de Coleta de Dados

O objetivo desta unidade é responder às perguntas: “Você considera que é importante a empresa ter conhecimento sobre quais dúvidas que o funcionário tem durante a realização de uma tarefa?” e “Você evitaria fazer as pesquisas de dúvidas na internet se soubesse que elas estão sendo coletadas para melhor entendimento das dificuldades e criação de cursos?”.

Considerando as respostas dos participantes, 75% consideram muito importante o fato de a empresa ter conhecimento sobre as suas dúvidas.

Dentre os participantes que consideraram importante o fato da empresa ter conhecimento sobre suas buscas, 66% não evitariam fazer buscas de dúvidas na internet mesmo sabendo que estas dúvidas seriam posteriormente analisadas pelo seu gestor.

Também nesta unidade foram respondidas as perguntas: “Você acredita que executaria de maneira mais eficaz seu trabalho se houvessem cursos com temas voltados para os projetos internos?” e “Você se considera apto para lecionar um curso dentro da empresa com o seguinte tema C# SQL DataReader?”.

De acordo com os participantes, 100% acreditam que o trabalho seria mais bem executado caso houvesse treinamentos dentro da empresa que auxiliassem na capacitação dos colaboradores. No entanto, nenhum participante considerou-se apto a aplicar um treinamento de um tema sugerido pelo ambiente. Este resultado mostra que os temas sugeridos são realmente a necessidade de aprendizagem e não temas nos quais os desenvolvedores já possuem conhecimento.

A próxima categoria a ser analisada é a Efetividade dos Temas. Os resultados das unidades dentro desta categoria são apresentados a seguir.

#### 1. Tema mais pesquisado

O objetivo desta unidade é responder à pergunta: “A seguinte busca foi realizada por um ou mais respondentes: “C# MVC Controller Action Result Json”. Os temas envolvidos nesta busca têm ligação com as tecnologias do seu dia a dia?”. A análise desta unidade foi realizada com base nas respostas de 50% dos participantes, que consideraram o tema apresentado na questão “C# MVC Controller Action Result Json” como tendo muita ligação com as tecnologias que trabalha no dia a dia. Este tema foi o mais pesquisado pelos participantes do experimento.

#### 2. Tema menos pesquisado

O objetivo desta unidade é responder à pergunta: “Neste experimento cerca de 200 pesquisas diferentes foram registradas. Dentre as pesquisas o seguinte resultado foi encontrado: “Index Online SQL Server”. Você considera



este assunto importante no seu dia a dia?”. O tema “Index Online SQL Server” foi um dos temas menos pesquisados pelos participantes. Com base em suas respostas, 75% dos participantes não consideraram o assunto importante para o dia a dia.

### 3. Relação entre palavras do Tema

O objetivo desta unidade é responder à pergunta: “Dentre os temas sugeridos para criação de cursos, foi sugerido “C# *Json Request Behaviour AllowGet MVC3*”. Você concorda que há relação entre todas as palavras do tema?”. O tema “C# *Json Request Behaviour AllowGet MVC3*” foi escolhido para ser avaliado pelos participantes se as palavras que compuseram o tema estão relacionadas. Dentre os participantes, 50% não souberam responder, enquanto que 25% afirmaram que as palavras estão relacionadas e 25% acreditam que não há relação entre as palavras.

Como esta unidade de avaliação não pode ser respondida de forma a obter um resultado positivo ou negativo quanto a relação entre as palavras do tema, uma pesquisa referente ao assunto foi realizada. Segundo o fórum online da Microsoft (MSDN, 2015), a classe “*JsonRequestBehaviour*” utilizada na linguagem de programação C#, especifica quando requisições HTTP GET são permitidas.

## 6.7 - Discussão

A fim de analisar a viabilidade da abordagem proposta para coleta de consultas realizadas em motores de busca para sugestão de temas de cursos, Identificador de Temas e Papéis (TIRONI et al., 2015), três hipóteses foram definidas e um experimento foi conduzido com objetivo de comprovar ou refutar as hipóteses.

Os resultados apresentaram que além de motores de busca serem utilizados para busca de necessidades de aprendizagem da equipe de desenvolvimento participante do experimento, existe um incentivo por parte de seus gerentes em melhorar a qualidade permitindo o uso da internet para este fim. Também é possível concluir que, apesar desta abordagem de coleta de *log*

do acesso à internet parecer intrusiva, a maioria dos participantes do experimento (75%) considera importante o fato da companhia saber a necessidade deste aprendizado.

Assim, a abordagem é aceita pelo grupo de desenvolvedores, principalmente quando se trabalha com o registro de *log* dos colaboradores, ou seja, não há a rejeição da possibilidade de adesão da abordagem de coleta de temas por meio das buscas na internet propostas neste trabalho. Portanto, com base neste experimento, a Hipótese H3 é verdadeira.

No que diz respeito aos temas sugeridos pelo ambiente, os desenvolvedores foram questionados sobre a necessidade de aprendizagem nos tópicos apresentados pelo mecanismo como os mais pesquisados. Para 75% dos participantes, os temas sugeridos pelo mecanismo foram considerados uma necessidade de aprendizagem para melhor execução das tarefas diárias. Desta forma, foi considerada verdadeira a Hipótese H1, a qual propõe que os temas sugeridos pelo mecanismo representam a necessidade de aprendizagem da maioria dos membros da equipe.

Da mesma forma, as respostas apontaram que as palavras dentro de cada tema sugerido estavam relacionadas umas com as outras, demonstrando que o mecanismo não mesclou assuntos que não estavam de alguma forma relacionados entre si. Também pode ser observado na execução do experimento, o algoritmo de clusterização, apesar de ter o limiar de similaridade ajustado, não apresentou temas repetidos, criando de forma coerente os grupos e efetuando a contagem de quantidade de buscas pelos mesmos temas. Portanto, de acordo com o experimento proposto, a hipótese H2 é verdadeira.

Assim, o experimento confirma que é viável a aplicação da abordagem de coleta de pesquisas realizadas em motores de busca. A abordagem mostrou-se eficaz na tarefa de descoberta de conhecimento que necessita ser construído dentro de organizações de desenvolvimento de software.

Segundo Menolli (MENOLLI, 2013), o especialista gerador do curso, deve conhecer não só o domínio, mas também as necessidades de quem vai usar as unidades de aprendizagem. Por essa razão, o experimento proposto neste Capítulo mostra resultados que indicam que o mecanismo Identificador

de Temas e Papéis (TIRONI et al., 2015) contribui com o Ambiente Colaborativo Semântico (MENOLLI; MALUCELLI; REINEHR, 2011).

## **6.8 - Considerações Finais**

Este capítulo apresentou um experimento realizado para verificar a viabilidade de implantar um mecanismo baseado na abordagem proposta no Capítulo 5, bem como analisar o comportamento dos algoritmos de forma a encontrar possíveis melhorias.

Por este trabalho possuir diferentes fases de execução do trabalho que se complementam, uma na qual faz a sugestão de temas, e a partir destes, outra fase auxilia na criação do esquema de aprendizagem, um segundo experimento para avaliação desta outra abordagem foi apresentado no Capítulo 7 a seguir.

## Capítulo 7 - **EXPERIMENTO PARA AVALIAÇÃO DO AMBIENTE**

Este capítulo tem como propósito principal apresentar o experimento conduzido com objetivo de avaliar o ambiente gerador de cursos implementado e descrito no Capítulo 5. O procedimento realizado neste experimento segue o modelo definido no Capítulo 6, portanto serão apresentados os objetivos, tipo, planejamento do experimento de avaliação do ambiente implementado para criação de cursos. Também foi exposto o método de execução do experimento, e por fim seus resultados e discussão.

### **7.1 - Introdução**

O uso de motores de busca dentro de empresas desenvolvedoras de software para as consultas de códigos fonte para reuso ou utilização como exemplo ou referência é algo que já está se tornando típico, com muitas empresas liberando os acessos a estas tecnologias em prol de facilitar o trabalho de desenvolvimento de software. Além disso, foi identificado neste trabalho um meio para reaproveitar a interação do desenvolvedor de software com o motor de busca, de forma a descobrir qual é a necessidade de aprendizagem da equipe.

Esta abordagem contribuiu para a criação de um ambiente que além de apresentar para um especialista quais têm sido os temas mais pesquisados pela equipe de desenvolvimento de software, também permite que a partir destes temas sejam elaborados cursos. Como apresentado no Capítulo 5, alicerçado em conceitos educacionais sobre elaboração de cursos, o ambiente proporciona que o curso seja gerado por etapas. Por fim, o ambiente estrutura o Esquema de Aprendizagem.

Assim sendo, um experimento foi planejado e executado para avaliar o ambiente implementado para criação de cursos e o Esquema de Aprendizagem gerado pelo ambiente. Diversos critérios foram estabelecidos, de maneira que os diferentes aspectos referentes a cada objeto avaliado possam ser analisados e objetivamente explorados.

## 7.2 - O experimento

O mecanismo Identificador de Temas e Papéis (TIRONI et a., 2015) foi implementado para coletar pesquisas realizadas em motores de busca como o Google, assim como o ambiente para Definição de Cursos. Neste Capítulo é apresentado o experimento para analisar como o ambiente definidor de cursos deve ser utilizado, quais os pontos relevantes apresentados pelo ambiente, como a ferramenta pode contribuir para a construção de competências, e como ocorre a integração entre o ambiente proposto neste trabalho e o Ambiente Colaborativo Semântico (MENOLLI; MALUCELLI; REINEHR, 2011).

Por se tratar de um ambiente que utiliza os temas gerados pelo Identificador de Temas e Papéis para criação de cursos, neste experimento foi proposto novamente a coleta de dados por parte de desenvolvedores de software. No entanto, este experimento tem o objetivo de utilizar os temas para a criação de um curso e do Esquema de Aprendizagem, que é objeto resultante deste trabalho.

## 7.3 - O método experimental

O presente experimento é caracterizado como um estudo experimental, assim como o experimento apresentado no Capítulo 6. Este é um experimento controlado em um objeto de estudo. Portanto, pode-se classificá-lo de acordo com Basili (BASILI, 1996) como:

- Estudo descritivo.
- *In vitro*.
- Realizado com especialistas;
- Estudo observacional;

## 7.4 - Descrição dos Objetivos

Este experimento tem dois principais objetivos.

1. Avaliar a abordagem de criação de cursos por etapas na visão de especialistas técnicos e de recursos humanos.
2. Gerar de forma semiautomática o Esquema de Aprendizagem e testar a execução do esquema diretamente no Ambiente Colaborativo Semântico proposto em Menolli, Reinehr e Malucelli (MENOLLI; REINEHR; MALUCELLI, 2013).

## **7.5 - Planejamento**

O planejamento foi dividido em quatro etapas. Cada etapa é descrita nas seções seguintes.

### **7.5.1. Definição das Hipóteses**

Para orientar este experimento, duas hipóteses foram estabelecidas com base nos objetivos deste experimento, que são descritas abaixo:

Hipótese (H1): Os especialistas consideraram que a abordagem para geração do Esquema de Aprendizagem contribui para o desenvolvimento de competências dentro da organização.

Hipótese (H2): O XML do Esquema de Aprendizagem gerado é executável e aceito pelo Ambiente Colaborativo Semântico proposto em Menolli, Malucelli e Reinehr (MENOLLI; MALUCELLI; REINEHR, 2011).

### **7.5.2. Seleção dos Participantes**

Este experimento foi dividido em duas etapas. Em razão da necessidade de dados reais para a criação de cursos, a primeira etapa, foi novamente utilizar a coleta de dados e resultados do Identificador de Temas e Papéis. Assim, na segunda etapa, foi possível utilizar dados não simulados para geração de cursos.

Portanto, para a primeira etapa, os mesmos critérios e instrumentos do experimento apresentado no Capítulo 6 foram utilizados na seleção dos

participantes. Dessa maneira, os participantes neste experimento foram dezoito colaboradores de empresas de desenvolvimento de software.

Foram coletadas em entrevista com dezoito os participantes seus dados para definição do perfil profissional de cada um. Desta forma, é possível definir o grupo como sendo desenvolvedores com experiência de dois a dez anos, atuando com as linguagens de programação C#, Delphi, Java e SAP.

Para a segunda etapa deste experimento, dois profissionais foram selecionados. Cada profissional selecionado possui um perfil. A necessidade destes perfis diferentes na avaliação se dá pelo motivo de que para a criação de cursos é necessário um profissional que tenha o conhecimento técnico em desenvolvimento de software para apontar os conteúdos deste curso, enquanto que, para a avaliação das competências que serão desenvolvidas no aprendiz, é necessário a análise de um especialista de Recursos Humanos. O primeiro, que foi denominado Especialista Técnico, um desenvolvedor de sistemas com seis anos de experiência com a linguagem de programação C# e tecnologias web como HTML, CSS e Javascript. Este especialista técnico foi selecionado por possuir conhecimento avançado nos temas pesquisados pelos participantes na primeira etapa deste experimento. O segundo profissional, denominado Especialista de Recursos Humanos, um colaborador do setor de Recursos Humanos com quinze anos de experiência. Este especialista foi selecionado por ter experiência com programas de desenvolvimento de competências dentro de organizações.

### **7.5.3. Descrição do Experimento**

Almeja-se avaliar se o ambiente proposto pode prover uma forma prática e dinâmica de criação de cursos e, conseqüentemente, a geração de Esquemas de Aprendizagem, favorecendo a atuação de um Especialista Técnico em compromisso com a aprendizagem organizacional.

Como denotado na seção 7.5.2, este experimento foi dividido em duas etapas. Na primeira etapa, dezoito colaboradores de empresas de desenvolvimento de software foram solicitados a ler um documento de requisitos de um software. Neste documento foi apresentado para os

participantes um projeto fictício de desenvolvimento de um sistema ERP. Após o entendimento da proposta de projeto de desenvolvimento, os participantes foram solicitados a levantar suas dúvidas referentes à:

- Regras de negócios (Cálculos, limitações do usuário, etc.);
- Padrões de Programação;
- Arquitetura do Sistema; e
- Tecnologias exigidas no desenvolvimento;

Posteriormente, cada participante foi solicitado a acessar um motor de busca, e pesquisar as respostas para todas as questões levantadas. O objetivo desta parte inicial do experimento é registrar todas as pesquisas que foram realizadas em um motor de busca as quais foram necessárias para sanar as dúvidas geradas pelos desenvolvedores.

Nesta etapa do experimento foram utilizados os mesmos critérios definidos no experimento apresentado no Capítulo 6. O motor de busca utilizado para coletar as dúvidas geradas pelos desenvolvedores foi o Google, por meio da URL de acesso [HTTP://www.google.com.br/?nord=1](http://www.google.com.br/?nord=1). O uso desta URL para acesso ao Google permitiu que as buscas fossem coletadas pelo servidor proxy.

Na segunda etapa deste experimento foi solicitado ao especialista técnico ler a documentação de requisitos do projeto fictício apresentado aos desenvolvedores. No entanto, a leitura da documentação foi com o objetivo de entender a necessidade da tarefa e confrontá-las com os temas sugeridos pelo Identificador de Temas e Papéis. Os temas apresentados pelo mecanismo foram a clusterização das pesquisas realizadas pelos desenvolvedores na primeira etapa do experimento.

O especialista técnico foi requisitado a escolher um dos temas sugeridos pelo mecanismo e iniciar a criação de um curso. Após a finalização da criação do curso, um Esquema de Aprendizagem foi gerado. Este Esquema de Aprendizagem foi testado no Ambiente Colaborativo Semântico (MENOLLI; MALUCELLI; REINEHR, 2011).

Do mesmo modo, o especialista de Recursos Humanos, foi solicitado a participar do experimento de forma observacional, também realizando a leitura do projeto fictício. No entanto, este especialista realizou comentários em relação à abordagem desenvolvida, as utilidades em um ambiente



organizacional, e de que formas o ambiente pode contribuir para aprendizagem organizacional. A participação deste especialista também foi importante no experimento para avaliar o curso definido pelo especialista técnico. O especialista de recursos humanos analisou o curso criado e como ele poderia contribuir com as competências dos aprendizes selecionados.

#### **7.5.4. Execução**

Para que a segunda etapa do experimento, que se refere à definição de um curso por parte de um especialista, apresentasse resultados relevantes, foi considerado realizar a criação de um projeto de desenvolvimento fictício. Neste projeto fictício, uma empresa do ramo de engenharia solicitou um sistema ERP para uma fábrica de software na qual o participante era um colaborador. Foram apresentados no projeto os Requisitos de Negócio e os Requisitos Técnicos para criação do sistema. O projeto fictício apresentado neste experimento consta no Apêndice B.

A partir da leitura projeto fictício apresentado no Apêndice B, o colaborador foi convidado a gerar suas dúvidas e encontrar as respostas por meio de buscas no Google. Ao final das pesquisas, foram identificadas 50 pesquisas realizadas referentes a dúvidas que foram geradas na leitura do projeto apresentado aos participantes.

Para coletar a informação sobre o usuário responsável pela busca, estes funcionários foram cadastrados no ambiente com suas habilidades específicas e os endereços IP que cada funcionário utiliza para navegação. Neste experimento não houve a necessidade de categorizar os participantes, pois após realizar as pesquisas não foi aplicado nenhum questionário.

Posteriormente, o mecanismo foi utilizado para recuperar os logs referentes às pesquisas e armazená-los no banco de dados. Após o processo de transformação dos dados, mineração de textos e clusterização, foram gerados a partir das 50 pesquisas, 20 clusters.

Desta forma a primeira etapa do experimento foi concluída. Na segunda etapa, o especialista técnico foi convidado a ler o projeto fictício

apresentado no Apêndice B e conhecendo a necessidade deste projeto, analisar quais eram as dúvidas dos colaboradores, e assim criar um curso.

O especialista técnico escolheu um tema dentre os 20 sugeridos. A partir deste tema o especialista foi capaz de seguir a abordagem proposta para Definição de Cursos apresentada no Capítulo 5, definindo inicialmente o título, objetivos e requisitos. Posteriormente, o especialista definiu novos temas que não apareceram nas pesquisas, e organizou os temas de forma hierárquica formando uma estrutura de um curso. Por fim, o ambiente sugeriu um orientador para o curso e como possíveis aprendizes, os participantes que haviam pesquisado por aquele tema. O especialista também incluiu novos aprendizes para o curso de acordo com o nível de habilidade de cada participante que foi identificado em entrevista inicial com os participantes.

O ambiente realizou a geração do Esquema de Aprendizagem contendo os dados informados pelo especialista. O Esquema de Aprendizagem foi primeiramente conferido visualmente, e também executado no Ambiente Colaborativo Semântico.

Ainda na segunda etapa do experimento, um especialista em Recursos Humanos foi convidado a analisar o ambiente criado, os temas sugeridos para criação de cursos, a documentação do projeto fictício e as habilidades dos participantes do experimento. A partir destas informações o especialista pôde contribuir com sugestões segundo a forma de aplicação da ferramenta em um ambiente empresarial, possíveis limitações e pontos positivos que contribuiriam para a aprendizagem organizacional e construção de competências.

## **7.6 - Resultados**

Nesta seção, os principais resultados após a análise da criação de um curso por parte de um especialista, bem como a análise descritiva de um especialista de Recursos Humanos são apresentadas.

### **7.6.1. Definição de Cursos**

A primeira etapa do processo de criação de cursos analisada foi a seleção do tema dentre os temas sugeridos pelo mecanismo. Posteriormente, foi avaliado o método de definição de cursos que o especialista seguiu e concluiu a geração do curso com a criação do Esquema de Aprendizagem. Por fim, foi analisado o Esquema de Aprendizagem gerado, inicialmente visualmente, com objetivo de verificar se a estrutura do esquema estava de acordo com o Learning Design (IMS LD, 2003) apresentado no Capítulo 2, e em um segundo momento, o Esquema de Aprendizagem foi executado no Ambiente Colaborativo Semântico (MENOLLI; MALUCELLI; REINEHR, 2011) para verificar se a estrutura gerada é aceita de forma a integrar os dois ambientes. Nestas etapas que são descritas a seguir, além das ações realizadas pelo especialista técnico, também são apresentados os comentários do especialista de Recursos Humanos sobre cada etapa.

### 1. Seleção do Tema

Apesar do mecanismo de sugestão de temas ter se mostrado efetivo na realização do experimento apresentado no Capítulo 6, foi necessário novamente utilizar este processo para gerar dados atuais e dentro do processo deste experimento. No entanto, o mecanismo ainda não havia sido experimentado no quesito seleção de temas.

Como apresentado no Capítulo 5, o mecanismo permite a escolha de mais de um tema para criação de cursos. O motivo disso é a maior facilidade na criação de cursos, uma vez que vários temas podem ser utilizados.

Neste experimento, o especialista selecionou apenas um tema para criação de cursos, sendo este “Framework Apache Cordova”. Segundo o especialista este tema foi selecionado por ser dentre todos os temas, o que ele possuía melhor conhecimento e também por apresentar um número de 6 pesquisas realizadas sobre o mesmo tema, ou seja, do total de dezoito participantes, um terço (1/3) não tinham conhecimento sobre o assunto.

Como houve ajustes no mecanismo, após a realização do experimento do Capítulo 6, com objetivo de realizar melhorias, o tema selecionado pelo especialista foi melhor clusterizado, e assim, não apresentou temas relacionados.

Apesar do maior número de participantes no segundo experimento, no primeiro experimento, os 4 participantes foram solicitados a executar em torno de 50 pesquisas, enquanto que neste experimento, os 18 participantes fizeram pesquisas apenas sobre dúvidas que foram levantadas durante a leitura do projeto fictício. Isto também foi identificado como um facilitador para a não geração de temas relacionados.

Uma base de dados, contendo as habilidades referentes a desenvolvimento de todos os participantes, foi utilizada para encontrar nos temas, tecnologias específicas informadas pelos desenvolvedores. Assim, ao selecionar o tema “Framework Apache Cordova”, o ambiente identificou, a partir da palavra “Cordova”, um possível orientador que tinha esta habilidade com nível de experiência alto, e também a tecnologia como sendo um requisito para participação do curso.

Neste processo de seleção dos temas, foram notados alguns pontos para os quais existem possíveis melhorias, como por exemplo, ao selecionar um tema e descobrir uma habilidade técnica deste tema, a partir das palavras informadas, tentar relacionar quais outras habilidades técnicas estão ligadas ao tema selecionado. Para execução desta melhoria seria necessário a criação de uma tabela de relacionamento entre as tecnologias.

Nesta etapa, houveram pontos positivos observados pelo especialista de recursos humanos. O primeiro ponto evidenciado é a possibilidade de entender melhor o perfil dos colaboradores dentro de uma organização. Aqueles desenvolvedores que continuamente têm dificuldades no dia a dia podem ser mais bem observados pelo departamento de recursos humanos que poderão incentivar a evolução de suas habilidades. Além de ajudar a identificar as necessidades reais de aprendizagem dos funcionários, pode evidenciar, a partir de um histórico de pesquisas de colaboradores em específicos se houve evolução em suas competências, o que pode auxiliar num possível programa de cargos e salários.

## 2. Definição do Curso

O processo seguinte foi a definição de um curso. Com o tema “Framework Apache Cordova” selecionado, o especialista acessou o ambiente para definição de cursos. Esta abordagem, como apresentada no Capítulo 5,

foi dividida em 4 etapas. A primeira etapa foi a de informações gerais do curso. Estas informações são importantes, pois serão apresentadas no Ambiente Colaborativo Semântico (MENOLLI; MALUCELLI; REINEHR, 2011) no momento da geração de Unidades de Aprendizagem e também por serem informações necessárias para completar o IMS Learning Design (IMS, 2003).

O especialista informou nesta etapa, o título do curso como sendo “Framework Apache Cordova”, o subtítulo do curso como “Programação Mobile”, o objetivo “Criar projetos utilizando o apache cordova e comunicar com banco de dados” e as palavras chave foram “Framework”, “Mobile”, “Apache” e “Cordova”. Também foram informados os dados a respeito dos pré-requisitos para participação do curso, como por exemplo “Tecnologias HTML 5, CSS 3, KnockoutJS e MongoDB”.

A partir da segunda etapa, além da criação de cursos de forma sequencial, o especialista teve o auxílio das dicas de criação de curso definidas pelo modelo SOI (MAYER, 1996), sendo informado sobre importantes ações a serem tomadas na seleção, organização e integração das informações de forma a obter maior aprendizado com o curso que estaria sendo criado.

A segunda etapa foi a seleção de temas relacionados e também criação de novos temas. O especialista criou 11 novos temas que foram utilizados para criação de cursos. Com o tema selecionado a ferramenta não foi capaz de encontrar outros temas que estivessem relacionados, devido ao número de pesquisas e aos ajustes realizados no mecanismo de identificação de temas e papéis. Nesta etapa o ambiente auxiliou o especialista com dicas para focar em informações importantes para os aprendizes e evitar informações que seriam irrelevantes para alcançar o objetivo do curso.

A etapa que se seguiu foi a de organização dos temas. O especialista organizou uma estrutura com temas do curso, de forma a obter uma hierarquia de apresentação dos assuntos. Nesta etapa, o ambiente auxiliou o especialista a organizar as informações de forma hierárquica e obter uma estrutura de apresentação de informações das mais simples para as mais complexas. Nesta mesma etapa, o ambiente apresentou dicas importantes para integração das informações, como por exemplo, onde deveriam estar os exemplos, e preparar o aprendiz para um curso de nível superior.

A última etapa da criação de cursos foi a escolha dos participantes. O tema selecionado permitiu que o ambiente sugerisse seis participantes, que foram os responsáveis pelas buscas. Dentre estes seis participantes, um foi sugerido como possível orientador, pois possui um nível de experiência considerado “Pleno” na habilidade “Cordova”. O especialista nesta etapa selecionou outros participantes para o curso, dado que devido as suas habilidades levantadas na entrevista inicial, alguns demonstraram ter pouco conhecimento sobre o assunto “Cordova”. Por fim, o especialista selecionou a opção de geração do Esquema de Aprendizagem. O ambiente gerou o esquema de aprendizagem e gravou em um arquivo no formato XML.

Foi possível obter importantes considerações a respeito do processo de criação de cursos por parte do especialista de Recursos Humanos. Foi verificado que existe a necessidade de atuação de um especialista técnico, conhecedor do assunto no qual o curso será criado, pois somente este especialista saberá relacionar os assuntos que deverão ser aprendidos. Foi observado que a atuação de um colaborador de recursos humanos na fase de criação de cursos será na análise de competências que o curso se propõe a criar e se estas habilidades foram realmente alcançadas pelos colaboradores.

### 3. Esquema de Aprendizagem

Ao final do experimento, um Esquema de Aprendizagem foi gerado, baseado no Esquema de Aprendizagem definido pelo IMS Learning Design (IMS, 2003), e adaptado por Menolli (MENOLLI, 2013). No processo de geração do Esquema, todas as estruturas necessárias para utilização e execução foram respeitadas, inclusive a forma como são criadas as identificações entre os temas.

Para avaliar o esquema de aprendizagem criado pelo ambiente, foram utilizadas duas propostas. A primeira foi a comparação com um esquema de aprendizagem gerado por uma outra ferramenta denominada Reload Dist e que foi utilizada por Menolli (MENOLLI, 2013) para criação de Unidades de Aprendizagem. Esta primeira avaliação mostrou que o esquema de aprendizagem criado pelo ambiente foi gerado de forma correta em suas estruturas, pois apresentou as mesmas *tags* de estrutura que o esquema de aprendizagem gerado pela ferramenta Reload Dist.

A segunda proposta de avaliação foi a execução do Esquema de Aprendizagem no Ambiente Colaborativo Semântico proposto por Menolli, Malucelli e Reinehr (MENOLLI; MALUCELLI; REINEHR, 2011). Neste processo, o esquema de aprendizagem se mostrou novamente correto, pois não houve nenhuma recusa no Ambiente Colaborativo Semântico. Com o esquema de aprendizagem executado, não foi possível localizar nenhum objeto de aprendizagem, pois estes deveriam ser criados anteriormente.

Por esta etapa do experimento ser técnica e voltada para a execução de um objeto gerado pelo ambiente, não houveram avaliações por parte do especialista de recursos humanos.

## **7.7 - Discussão**

A fim de avaliar a abordagem proposta para criação de cursos, duas hipóteses foram definidas e um experimento foi conduzido com o objetivo de comprovar ou refutar as hipóteses.

Os especialistas, tanto o técnico quanto o de recursos humanos, consideraram que a abordagem para definição de cursos, contribui para o desenvolvimento de competências dentro de uma organização. Como afirmado por Paquette (PAQUETTE, 2002), uma competência é definida pela combinação de conhecimentos de saber, habilidades e atitudes. Este trabalho procura a partir de necessidades de aprendizagem permitir que a organização construa o saber de seus colaboradores exatamente nas habilidades que precisam ser construídas. Por esse motivo, a hipótese H1 foi considerada verdadeira.

No que diz respeito ao Esquema de Aprendizagem, que é um artefato originado da abordagem de criação cursos, as suas estruturas foram geradas pelo ambiente de forma correta, o que resultou na execução correta. Esse fato fornece subsídio para responder a hipótese H2 como verdadeira.

## **7.8 - Considerações Finais**

Este Capítulo apresentou o planejamento e execução de um experimento para avaliar o ambiente implementado e adquirir elementos suficientes para responder as hipóteses definidas no início deste experimento.

O experimento foi realizado seguindo o método experimental, utilizando um estudo em duas etapas: participantes desenvolvedores de software que realizaram buscas que auxiliaram a preencher uma base de dados para a posterior criação de cursos, e participantes considerados especialistas atuando na criação de um curso e avaliação do ambiente sob uma ótica empresarial.

Para avaliar o ambiente, diversos critérios foram definidos, explorando o uso do ambiente por um especialista técnico e um especialista de recursos humanos.

Após realizar o experimento e a avaliar o ambiente, as hipóteses do experimento foram analisadas e respondidas, tomando como base não apenas os dados coletados neste experimento, mas também no experimento anterior e no referencial teórico.



## Capítulo 8 - **CONCLUSÕES**

Após apresentar os resultados obtidos com esta pesquisa, este capítulo posiciona sua relevância, demonstrando as principais contribuições, limitações e perspectivas de trabalhos futuros.

### **8.1 - Relevância do Estudo**

Como apresentado no Capítulo 1 deste trabalho, é considerado que mesmo um especialista capacitado, que tenha conhecimento suficiente para gerar unidades de aprendizagem que auxiliem o funcionário em seu aprendizado de forma eficaz, tem um trabalho complexo quando se diz respeito a conhecer qual a necessidade de aprendizagem dos colaboradores da empresa, não somente para crescimento deste funcionário, mas que este conhecimento possa resultar na evolução de competências em habilidades.

Tendo em vista esta direção do problema, foi proposto um mecanismo sustentado por conceitos da aprendizagem organizacional e teorias pedagógicas com objetivo de contribuir com o especialista na composição destes Esquemas de Aprendizagem. Este mecanismo é apresentado como uma forma de semiautomatizar a criação destes esquemas a partir de pesquisas realizadas em motores de busca, que serão classificados em forma de temas, e posteriormente utilizados na definição de cursos.

### **8.2 - Contribuições da pesquisa**

Apesar de serem avaliados separadamente, o mecanismo Identificador de Temas e Papéis e o Definidor de Cursos se completam e são partes de um único ambiente. Com isso, são vislumbradas quatro contribuições principais com a conclusão do trabalho proposto:

A primeira contribuição desta pesquisa foi capturar as pesquisas realizadas em motores de busca, com intuito de catalogar temas e, de acordo com o conhecimento dos colaboradores da empresa, sugerir temas de cursos que devem ser gerados. Como analisado no experimento apresentado no

Capítulo 6, este mecanismo foi considerado viável para aplicação em um ambiente empresarial.

A segunda contribuição desta pesquisa é a criação de um ambiente semiautomática baseada em conceitos da área pedagógica para um problema identificado que é a geração de cursos. Como observado na implementação do ambiente, o especialista responsável em criar cursos dentro de uma organização normalmente é uma pessoa que não possui conhecimentos em elaboração de cursos. Por isso, este ambiente apresenta a vantagem de, baseado na teoria de elaboração de Reigeluth (REIGELUTH, 1999), com objetivo de desenvolvimento de habilidades dos colaboradores e utilizando-se de técnicas apresentadas pelo modelo de design instrucional SOI (MAYER, 1996), permite que o especialista seja amparado no momento da criação de cursos.

A terceira contribuição identificada nesta pesquisa foi a implementação de um componente capaz de ser integrado ao ambiente colaborativo semântico apresentado por Menolli, Malucelli e Reinehr (MENOLLI; MALUCELLI; REINEHR, 2011), que seja eficaz no trabalho de contribuir com o especialista na geração de Esquemas de Aprendizagem, de forma a tornar a aprendizagem organizacional mais efetiva.

E por fim, a última contribuição identificada foi a publicação de um artigo no Congresso Internacional de Engenharia de Software e Engenharia de Aprendizagem com o título “*An approach to identify relevant subjects for supporting the learning scheme creation task*” em 2015 (TIRONI et al., 2015).

### **8.3 - Limitações**

A principal limitação da pesquisa está concentrada na forma como o motor de busca deve ser utilizado em uma rede corporativa. No mecanismo foi utilizado o motor de busca semântico para base de sua implementação. Para que a busca realizada seja identificada por um servidor Proxy, é necessário que este motor de busca não faça as buscas utilizando o protocolo seguro de encriptação. Assim, primeiramente não se pode garantir que o uso de um motor de busca que não apresente esta característica acarrete os mesmos resultados, uma vez que o motor de busca testado no mecanismo tem

características que outros motores de busca não apresentam. Ainda com relação ao motor de busca, para cada novo motor de busca utilizado na coleta de dados de pesquisa, foi necessário definir e implementar um novo tratamento para extrair os dados relevantes para definição dos temas.

#### **8.4 - Trabalhos Futuros**

Por fim, são apresentadas como perspectivas de trabalhos futuros: (i) coletar dados referentes as pesquisas realizadas, não apenas utilizando de motores de busca, mas também, de páginas acessadas pelo desenvolvedor no momento da consulta; (ii) utilizar um algoritmo para adaptação do limiar de similaridade no momento da clusterização dos temas. Este algoritmo poderá ajudar a comparar grupos de palavras encontrando os grupos similares independentes da quantidade de palavras informadas na consulta em motores de busca.

## REFERÊNCIAS BIBLIOGRÁFICAS

(ADL, 2004) Advanced Distributed Learning. **Scorm 3rd Edition Content Aggregation Model**. 2004. Disponível em <<http://www.adlnet.gov>> 2004. Acesso em: 15 de Maio 2011

(ADOMAVICIUS; TUZHILIN, 2005) ADOMAVICIUS, G.; TUZHILIN, A. **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, v. 17, n. 6, Junho, 2005.

(AGGARWAL; ZHAI, 2012) AGGARWAL, C.C.; ZHAI, C.X. **Mining Text Data**. Springer Science & Business Media, 2012.

(AGRAWAL et al., 1998) AGRAWAL, R.; GEHRKE, J.; GUNOPULOS, D.; RAGHAVAN, P. **Automatic Subspace CLustering of High Dimensional Data for Data Mining Applications**. IMB Almaden Research Center, 1998.

(ALE et al., 2014) ALE, M. A.; TOLEDO, C. M.; CHIOTTI, O., GALLI, M. R. **A conceptual model and technological support for organizational knowledge management**. Special Issue on Systems Development by Means of Semantic Technologies, v. 95, p. 73-92, 2014.

(ALI; PASCOE; WARNE, 2002) ALI, I. M.; PASCOE, C.; WARNE, L. **Interations of organization culture and collaboration in working and learning**. Education Technology and Society, v. 5, n. 2, p. 60-69, 2002.

(ALVARES; GARCIA; FERRAZ, 2005) ALVARES, R.; GARCIA, A.; FERRAZ, I. **A Stemming Algorithm for the Brazilian Portuguese Language**. IN: Progress in Artificial Intelligence, v. 3808, p. 693-701, 2005.

(ALVESSON, 2000) ALVESSON, M. **Social identity and the problem of loyalty in knowledge-intensive companies.** Journal of Management Studies, v. 37, n. 8, p. 1101-1123, 2000.

(AMORIM et al., 2006) AMORIM, R. R.; LAMA, M.; SÁNCHEZ, E.; RIEIRA, A.; VILA, X. A. **A Learning Design Ontology based on the IMS Specification.** Educational Technology & Society, v. 9, p. 38-57, 2006.

(ANDRADE et al., 2012) ANDRADE, J.; ARES, J.; MARTÍNEZ, M.; PAZOS, J.; RODRÍGUEZ, S.; ROMERA, J.; SUÁREZ, S. **An architectural model for software testing lesson learned systems.** Information and Software Technology, 2012.

(ANTONIOL et al., 2000) ANTONIOL, G.; CANFORA, G.; GERARDO, C.; LUCIA, A. **Information Retrieval models for recovering traceability links between code and documentation.** In Proc. of IEEE Intl. Conf. on Software Maintenance, San Jose, CA, USA, 2000.

(ARGYRIS; SCHÖN, 1996) ARGYRIS, C.; SCHÖN, D. **Organizational Learning II: Theory, Method, and Practice.** Reading, MA, USA Cambridge, Uk: Addison Wesley, 1996.

(BARDIN, 1977) BARDIN. L. **L'analyse de contenu.** Presses universitaires de France, France, 1977.

(BASILI, 1996) BASILI V. **The Role of Experimentation in Software Engineering: Past, Present, Future.** IN : PROC. OF THE 18<sup>TH</sup> INTERNATIONAL CONFERENCE ON SOFTWARE ENGINEERING, 1(2), 1996, p. 133-164.

(BENNETT, et al., 2012) BENNETT, S.; BISHOP, A.; DALGARNO, B.; WAYCOTT, J.; KENNEDY, G. **Implementing Web 2.0 technologies in higher education: A collective case study.** Expert Systems with Applications, v. 59, n. 2, p. 524–534, 2012.

(BERTCHOLD et al., 1997) BERCHOLD, S.; KEIM, D.; KRIEGEL, H. **A cost model for nearest neighbor search in high-dimensional data space**. In Proceedings of the 16<sup>th</sup> ACM Symposium on Principles of Database Systems (PODS 1997), Tucson, Arizona, USA, Maio, 1997.

(BERNERS-LEE; HENDLER; LASSILA, 2001) BERNERS-LEE, T.; HENDLER, J.; LASSILA, O. **The Semantic Web**. Scientific American, v. 284, n. 5, p. 34–43, 2001.

(BIASUTTI; EL-DEGHAIIDY, 2012) BIASUTTI, M.; EL-DEGHAIIDY, H. **Using Wiki in teacher education: Impact on knowledge management processes and student satisfaction**. Computers & Education, v.59, p861-872, 2012.

(BOH, 2007) BOH, W. **Mechanisms for sharing knowledge in project-based organizations**. Information and Organization, v. 17, n.1, p. 27-58, 2007.

(BOMARIUS et al., 2000) BOMARIUS, F.; ALTHOFF, K.-D.; MÜLLER, W. **Knowledge management for learning software organizations**. Information Systems Frontiers, v. 2, n. 4, p. 349-367, 2000.

(BRABHAM, 2008) BRABHAM, D. **Crowdsourcing as a Model for Problem Solving, an Introduction and Cases**. The international Journal of Research into New Media Technologies, v. 14, 2008.

(BRASIL, 2007) Ministério da educação. Secretaria de Educação a Distância. **Objetos de aprendizagem: uma proposta de recursos pedagógicos**. Brasília: Mec, pp. 154, 2007.

(BRIN; PAGE, 1998) BRIN, S.; PAGE, L. **The anatomy of a large-scale**. Computer Networks, v. 30, pp. 107 – 117, 1998.

(BRITO et al., 2002) BRITO, S. R.; TAVARES, O. L.; MENEZES, C. S. **MEDIADOR - Um ambiente para aprendizagem orientada a projetos com**

**suporte inteligente à mediação.** IN: SIMPÓSIO BRASILEIRO DE INFORMÁTICA NA EDUCAÇÃO, 1, 2002. Anais...2002, p.116.

(BRUNER, 1966) Bruner, J. S. **Toward a theory of instruction**, Cambridge, Mass.: Belkapp Press, 1966.

(BURKE, 2002) BURKE, R. **Hybrid Recommender Systems: Survey and Experiments.** User Modeling and User-Adapted Interaction, v. 12, n. 4, pp. 331-370, 2002.

(CAPUANO; MIRANDA; ORCIUOLI, 2009) CAPUANO, N.; MIRANDA, S.; ORCIUOLI, F. **IWT: A Semantic Web-based Educational System.** IN: WORKSHOP OF THE WORKING GROUP ON AI & E-LEARNING , 4, 2009. Anais..., 2009, p. 11-16.

(CARRERAS et al., 2011) CARRERAS, M.; MARIN, P. M.; BERNAL, J. B; ALCARAZ, J. C, J.; MARTINEZ, G. P.; GOMEZ, G. S. **Towards a Semantic-Aware Collaborative Working Environment.** Computing and Informatics, v. 30, n. 1, p. 7-30, 2001.

(CHAPMAN, 2009) CHAPMAN, S. **SimMetrics.** Disponível em: <http://www.dcs.shef.ac.uk/~sam/stringmetrics.html>, 2014.

(CHEN, 2005) CHEN, Y. **Information Valuation for Information Lifecycle Management**, 2nd International Conference on Autonomic Computing, Seattle, WA, 2005.

(CIO Association, 2014) CIO Association. **Focusing on the fundamentals: Enterprise IT trends and investments.** EY, 2014.

(COOLEY et al., 1999) COOLEY, R.; MOBASHER, B.; SRIVASTAVA, J. **Web Mining: Information and Pattern Discovery on the World Wide Web.** Proc. 9<sup>th</sup> IEEE Int'l Conf. on Tools with Artificial Intellingence, 1999.

(COSTA et al., 2008) COSTA, D. P.; DOURADO, P.; PEREIRA, A.; SANTANCHÈ, A. **O Desenvolvimento de Jogos Baseado em OA para Learning**. IN: XIX SIMPÓSIO BRASILEIRO DE INFORMÁTICA NA EDUCAÇÃO - SBIE 2008, 2008, Fortaleza - CE. Anais..., 2008. p. 1-10.

(CUBRANIC et al., 2004) CUBRANIC, D.; MURPHY, G. C.; SINGER, J.; BOOTH, K. S. **Learning from project history: a case study for software development**. IN: ACM CONFERENCE ON COMPUTER SUPPORTED COOPERATIVE WORK. Anais...2004, p.82–91.

(DAVANZO et al., 2008) DAVANZO, E.; ELIA, A.; LIETO, A.; PREZIOSI, R. **Where does text mining meet knowledge management? A case study**. The university of Salerno, Fisciano (SA), Italy.

(DAVENPORT; PRUSAK, 1998) DAVENPORT, Thomas H.; PRUSAK, Lawrence. **Working Knowledge: How Organizations Manage What They Know**. Boston, MA, USA: Harvard Business School Press, 1998.

(DAVENPORT; PRUSAK, 2003) DAVENPORT, Thomas H.; PRUSAK, Lawrence. **Conhecimento empresarial: Como as organizações gerenciam o seu capital intelectual**. 14 ed. Rio de Janeiro: Futura, 2003. 243 p. Disponível em: <http://books.google.com.br/books?id=p4n9TZ9J3bAC> Acesso em: 11 Jan. 2014.

(DEERWESTER et al., 1990) DEERWESTER, S.; DUMAIS, S.; FURNAS, G.; LANDAUER, T.; HARSHMAN, R. **Indexing by Latent Semantic Analysis**. Journal of the American Society of Information Science, vol. 41, no. 6, 391-407.

(DIAS et al., 2009) DIAS, C. C.; KEMCZINSKI, A.; LUCENA, S. V.; FERLIN, J.; HOUNSELL, M. S. **Padrões abertos: aplicabilidade em Objetos de Aprendizagem (OAs)**. IN: SIMPÓSIO BRASILEIRO DE INFORMÁTICA NA EDUCAÇÃO, 2009. Anais..., 2009.



(DRUCKER, 1993) DRUCKER, P. F. **Post-capitalist society**. New York: Harper Business, 1993.

(EKOKO; ROBILLARD, 2012) EKOKO, E.; ROBILLARD, M.P. **Asking and Answering Questions about Unfamiliar APIs: An exploratory study**. In: Proceedings of the ACM/IEEE International Conference On Software Engineering, pp. 266-276, 2012.

(ERTOZ; STEINBACH; KUMAR, 2001) ERTOZ, L.; STEINBACH, M.; KUMAR, V. Finding Topics in Collections of Documents: **A Shared Nearest Neighbor Approach**. In Proceeding of Text Mining Workshop, First International SIAM Data Mining Conference, Chicago, IL, 2001.

(FALBO, 1998) FALBO, R. A. **Integração De Conhecimento Em Um Ambiente De Desenvolvimento De Software**. 1998, Tese (Doutorado), Programa De Pós-Graduação De Engenharia – Coppe, Universidade Federal do Rio de Janeiro (UFRJ), Rio de Janeiro, 1998.

(FAYYAD; PIATETSKY-SHAPIRO; SMITH, 1996) Fayyad, U.; PIATETSKY-SHAPIRO, G.; SMITH, P. **From Data Mining to Knowledge Discovery in Databases**. American Association for Artificial Intelligence, p. 37 - 54, 1996.

(FELDMAN et al., 1995) FELDMAN, R.; DAGAN, I. Knowledge Discovery in Textual Databases (KDT). IN: KDD-95 Proceedings, 1995.

(FELDMANN; MÜNCH; VORWIEGER, 1998) FELDMANN, R. L.; MÜNCH, J.; VORWIEGER, S. **Towards Goal-Oriented Organizational Learning: Representing and Maintaining Knowledge in an Experience Base**. IN: INTERNATIONAL CONFERENCE ON SOFTWARE ENGINEERING AND KNOWLEDGE ENGINEERING, 10, 1998. Anais..., 1998, p. 236-245.

(GAO; CHANG; HAN, 2005) Gao, L., Chang, E., HAN., S., 2005. **Powerful Tool to Expand Business Intelligence: Text Mining**. World Academy of Science, Engineering and Technology, 8, 110-115

(GARVIN, 1993) GARVIN, D. A. **Building a learning organization**. Harvard Business Review, v. 71, n. 4, p. 78-91, 1993.

(GIL, 2002) GIL, A. C. **Como elaborar projetos de pesquisa**. 4. ed. São Paulo: Atlas, 2006. 175 p.

(GIL; RATNAKAR, 2002) GIL, Y.; RATNAKAR, V. **A Comparison of (Semantic) Markup Languages**. IN: INTERNATIONAL FLAIRS CONFERENCE, 2002, PENSACOLA BEACH, FLORIDA, USA. **Anais....**, 2002.

(GOLDBERG et al., 1992) GOLDBERG, D.; NICHOLS, D.; OKI, B. M.; TERRY, D. **Using Collaborative Filtering to Weave an Information Tapestry**. Communications of the ACM, New York, v. 35, n. 12, pp. 61-70, dezembro, 1992.

(GONZÁLEZ; ROSA; MONTANER, 2007) GONZÁLES, G.; DE LA ROSA, J.L.; MONTANER, M. **Embedding Emotional Context in Recommender Systems**. In The 20<sup>th</sup> International Florida Artificial Intelligence Research Society Conference-FLAIRS, Key West, Florida.

(GOOGLE, 2015) GOOGLE. **How Search Works**. Disponível em: <http://www.google.com/insidesearch/howsearchworks/thestory/> - Acesso em: 2015.

(HEARST, 1999) HEARST, M. **Untangling Text Data Mining**, 37th Annual Meeting of the Association for Computational Linguistics, College Park, MD, 1999.

(HENNINGER, 2001) HENNINGER, S. **Turning development standards into repositories of experiences**. Software Process: Improvement and Practice, Wiley Online Library, v. 6, n. 3, p. 141–155, 2001.

(HEQUET, 1995) HEQUET, M. **Not paid enough? You're not alone**. Training, 32 (11), 44-55, 1995.

(HOLZ; KÖNNECKER; MAURER, 2001) HOLZ, H., KÖNNECKER, A., MAURER, F. **Task-specific knowledge management in a process-centred see**. Advances in Learning Software Organizations, Springer, 2001.

(HOTH0; NUNRBERGER; PAAB, 2005) HOTH0, A.; NURNBERGER, A.; PAAB, G. **A Brief Survey of Text Mining**. LDV Forum, 2005.

(HUANG, 2008) HUANG, A. **Similarity Measures for Text Document Clustering**. New Zealand Computer Science Research Student Conference, p 49 -56, 2008.

(IEEE, 2002) IEEE Computer Society. **Draft Standard for Learning Object Metadata**. Disponível em: [http://ltsc.ieee.org/wg12/files/LOM\\_1484\\_12\\_1\\_v1\\_Final\\_Draft.pdf](http://ltsc.ieee.org/wg12/files/LOM_1484_12_1_v1_Final_Draft.pdf). Acesso em 21 mar, 2014. New York: IEEE, 2002.

(IEEE, 2004) IEEE Computer Society. **Guide to the Software Engineering Body of Knowledge (SWEBOK)**. Disponível em: <http://www.computer.org/portal/web/swebok/htmlformat>>. Acesso em: 10 ago. 2010. New York: IEEE, 2004.

(IMS, 2003) IMS Global Learning Consortium Inc. **Learning Design Specification**. 2003. Disponível em: <http://www.imsglobal.org/learningdesign/index.html>. Acesso em: 21 mar. 2014.

(JARRAHI; KANGAVARI, 2012) JARRAHI, A.; KANGAVARI, M. **An Architecture for Context-Aware Knowledge Flow Management Systems**. International Journal of Computer Science Issues, v. 9, n. 3, p. 40-51, 2012.

(JIN; CORDY, 2005) JIN, D.; CORDY, J. **Ontology-Based Software Analysis and Reengineering Tool Integration: The OASIS Service-Sharing Methodology**. In 21st IEEE International Conference on Software Maintenance (ICMS), 2005.

(KIMMERLE, 2010) KIMMERLE, J. **Social Software and Knowledge Building: Supporting Co-Evolution of Individual and Collective Knowledge.** IN: PROCEEDINGS OF THE 9TH INTERNATIONAL CONFERENCE OF THE LEARNING SCIENCES. Anais..., 2010, p. 9-16.

(KIMMERLE; MOSKALIUK; CRESS, 2009) KIMMERLE, J.; MOSKALIUK, J.; CRESS, U. **Learning and Knowledge Building with Social Software.** IN: INTERNATIONAL CONFERENCE ON COMPUTER SUPPORTED COLLABORATIVE LEARNING, 2009. Anais..., 2009, p. 459-468.

(KINTSCH, 1988) KINTSCH, W. The use of knowledge in discourse processing: A construction-integration model. *Psychological Review*, 95, 163-182, 1988.

(KOCH, 2002) KOCH, M. **Interoperable Community Platforms and Identity Management in the University Domain.** *International Journal on Media Management*, v. 4, n. 1, p. 21-30, 2002.

(KODRATOFF, 1999) KODRATOFF, Y. Knowledge Discovery in Texts: **A Definition, and Applications.** IN: 11th International Symposium on Foundations of Intelligent Systems, 1999.

(KOPER; OLIVIER, 2004) KOPER, R.; OLIVIER, B. **Representing the Learning Design of Units of Learning.** *Educational Technology & Society*, 7, 97-111.

(KOSALA; BLOCKELL, 2000) KOSALA, R.; BLOCKEEL, H. **Web Mining Research: A Survey.** In *SIGKDD Explorations*, v. 2, i. 1, Julho, 2000.

(KOWALSKI, 2007) KOWALSKI, G. *Information Retrieval Systems: Theory and Implementation.* Kluwer Academic Publishers, v. 3, 1999.

(KRONE; SYVÄJÄRVI; STENVALL, 2009) KRONE, O.; SYVÄJÄRVI, A.; STENVALL, J. **Knowledge Integration for Enterprise Resources Planning**

**Application Design.** Knowledge and Process Management, v. 16, n. 1, p. 1-12, 2009.

(LE BOTERF, 1995) LE BOTERF, G. **De la compétence – essai sur um attracteur étrange.** In: Les éditions d'organisations. Paris: Quatrième Tirage, 1995.

(LESKOVEC; RAJAMARAN; ULLMAN, 2014) LESKOVEC, J.; RAJAMARAN, A.; ULLMAN, J. **Mining of Massive Datasets.** Stanford. 2014.

(LEVENE, 2011) LEVENE, M. **An introduction to search engines and web navigation.** John Wiley & Sons, 2011.

(LEVINE; MONARCH, 1998) LEVINE, L.; MONARCH, I. **Collaborative Technology in the Learning Organization: Integrating Process with Information Flow, Access, and Interpretation.** IN: IEEE INTERNATIONAL CONFERENCE ON SYSTEMS SCIENCES, 31., 1998, Hawaii. Anais..., p. 444-461, 1998.

(LI; RILLING, 2008) WITTE, R.; Li, Q.; ZHANG, Y.; RILLING, J. **Text mining and Software Engineering: An Integrated Source Code and Document Analysis Approach.** IET Software Journal, v. 2, n. 1, 2008.

(LICHTNOW; GARIN; PALAZZO, 2009) LICHTNOW, D. ; GARIN, R. S.; PALAZZO, L. A. **O Uso de Técnicas de Recomendação em um Sistema para Apoio à Aprendizagem Colaborativa.** Revista Brasileira de Informática na Educação, v. 14, n. 3, 2009.

(LOPS; GEMMIS; SEMERARO, 2011) LOPS, P.; GEMMIS, M.; SEMERARO, G. **Content-based Recommender Systems: State of the Art and Trends,** 2011.

(LUGO, 2004) LUGO, A. G. **Um Modelo de Sistemas Multiagentes para Partilha do Conhecimento Utilizando Redes Sociais Comunitárias.** 2004,

227 Tese (Doutorado) Escola Politécnica, Universidade de São Paulo (USP), São Paulo, 2004.

(LUO; CAO, 2009) LUO, Y.; CAO, F. **Web Knowledge Management System Based on User Model**, ICIE '09, 1, 552-556, 2009.

(MACHADO et al. 1999) MACHADO, M.; SANTOS, F.; WERNER, C.; BORGES, M. **Uma infra-estrutura de Apoio à Aquisição Cooperativa de Conhecimento em Engenharia de Domínio**. IN: SIMPÓSIO BRASILEIRO DE ENGENHARIA DE SOFTWARE (SBES'99), FLORIANOPOLIS (BRASIL), 1999. Anais... , 1999.

(MANNING, RAGHAVAN, SCHUTZE, 2008) MANNING, C. D., RAGHAVAN, P., SCHUTZE, H., 2008. **An Introduction to Information Retrieval**, Cambridge University Press, 421-442.

(MAYER, 1996) Mayer, R. H. (1999). **Designing instruction for constructivist learning**. In C. M. Reigeluth, (Eds.), In C. M. Reigeluth, (Ed), *Instructional-design theories and models: A new paradigm of instructional theory*, Volume II. pp,141-160. Mahwah, NJ: Lawrence Erlbaum Associates.

(MENDONÇA NETO et al., 2001) MENDONÇA NETO, M. G. DE; SEAMAN, C. B.; BASILI, V.; KIM, Y. M. **A prototype experience management system for a software consulting organization**. IN: INTERNATIONAL CONF. SOFTWARE ENG. AND KNOWLEDGE, CCITeseer, 2001. Anais...2001, p. 29–36.

(MENOLLI; MALUCELLI; REINEHR, 2011) MENOLLI, A. L.; MALUCELLI, A.; REINEHR, S. **Towards a Semantic Social Collaborative Environment for Organizational Learning**. IN: INTERNATIONAL CONFERENCE ON INFORMATION TECHNOLOGY AND APPLICATIONS (ICITA), 7, 2011, SIDNEY, AUSTRALIA. Anais..., 2011, p. 65-70.

(MENOLLI; REINEHR; MALUCELLI, 2013a) MENOLLI, A. L.; REINEHR, S.; MALUCELLI, A. **Organizational Learning Applied to Software Engineering:**

**A Systematic Review.** International Journal of Software Engineering and Knowledge Engineering. 2013.

(MENOLLI; REINEHR; MALUCELLI, 2013b) MENOLLI, André Luís Andrade; REINEHR, Sheila; MALUCELLI, Andreia. **Improving Organizational Learning: Defining Units of Learning from Social Tools.** Informatics in Education, v. 12, p. 273-290, 2013.

(MENOLLI, 2013) MENOLLI, André L. A., **Ambiente Colaborativo Social Semântico Voltado à Aprendizagem Organizacional Para Empresas de Desenvolvimento de Software,** Tese (Doutorado em Engenharia de Software). Programa de Pós-Graduação em Informática da Pontifícia Universidade Católica do Paraná, p. 278, 2013.

(MENOLLI et al., 2014) MENOLLI A.; PINTO S.; REINEHR S.; MALUCELLI A. **Unit of Organizational Learning Ontology based on LOM Standard and IMS Learning Design.** IN: Formal Ontology in Information Systems (FOIS), v. 267, p. 396-405, 2014.

(MORAES; GALIAZZI, 2006) MORAES R., GALIAZZI, M. C. **Discursive textual analysis: a multiple face reconstrutive process.** Ciência & Educação, v. 12, n. 1, p.117-128, 2006.

(MSDN, 2015) MSDN. **JsonRequestBehaviour** Enumeration. Disponível em: [https://msdn.microsoft.com/en-us/library/system.web.mvc.jsonrequestbehavior\(v=vs.118\).aspx](https://msdn.microsoft.com/en-us/library/system.web.mvc.jsonrequestbehavior(v=vs.118).aspx). Acessado em 10 de agosto de 2015.

(NAHM; MOONEY, 2002) NAHM, U. Y.; MOONEY, R. J. **Mining softmatching association rules.** In Proceedings of the Eleventh International Conference on Information and Knowledge Management (CIKM-2002), 681–683, 2002.

(NEVES, 2011) NEVES, E. O. **Aprendizagem Organizacional: considerações sobre metodologias de promoção e desenvolvimento.** Revista da Faculdade de Administração e Economia, v. 3, PP. 2 – 16, 2011.

(NETMARKETSHARE, 2015) **NETMARKETSHARE.** Disponível em: <https://netmarketshare.com/> - Acesso em 2015.

(NEVIS; DI BELLA; GOULD, 1995) NEVIS, E. C.; DI BELLA, A.; GOULD, J. M. **Understanding organizations as learning systems.** Sloan Management Review, v. 36, n.2, p. 73-85, 1995.

(NONAKA; KONNO, 1998) NONAKA, I.; KONNO, N. **The concept of “Ba”:** **Building a Foundation for Knowledge Creation,** California Management Review, Usa, v. 40, n. 3, p. 40-54, 1998.

(NONAKA; TAKEUCHI, 1997) NONAKA, I; TAKEUCHI, H. **Criação de conhecimento na empresa: como as empresas japonesas geram a dinâmica da inovação.** Rio de Janeiro: Campus, 358 p., 1997.

(PAQUETTE, 2001) PAQUETTE, G. **Modélisation des connaissances et des competences: Un langage graphique pour conceviur et apprendre,** 2001.

(PETRI; OLIVEIRA; FONTOURA, 2013) PETRI, G.; OLIVEIRA, T. B. de; FONTOURA, L. M. **Perspectivas do Uso da Internet em Empresas de Desenvolvimento de Softwares.** Anais EATI, p. 117-123, 2013.

(PÊCHEUX, 1995). PÊCHEUX, Michel. **Semântica e Discurso. Uma crítica à afirmação do óbvio.** Trad. Eni P. Orlandi (et al.). 2. ed. Campinas: Ed. da Unicamp, 1995.

(POLSANI, 2004) POLSANI, P. R. **Use and abuse of reusable learning objects.** Journal of Digital Information, v. 3, n.4, 2004.



(POPPER; LIPSHITZ, 2000) POPPER, M.; LIPSHITZ, R. **Organizational Learning: Mechanisms, culture, & feasibility**. Management Learning, v. 31, n. 2, p. 181-196, 2000.

(PORTER, 1980) PORTER, M. **Competitive strategy: techniques for analyzing industries and competitors**. New York: Free Press, 1980.

(PTSTEMMER, 2015) PTSTEMMER. **A Stemming toolkit for the Portuguese language**. Disponível em: <https://code.google.com/p/ptstemmer/>. Acesso em: 2015.

(RAMOS, BRÄSCHER, 2009) RAMOS, H. S. C.; BRÄSCHER, M. **Aplicação da descoberta de conhecimento em textos para apoio à construção de indicadores infométricos para a área de C&T**. Ci. Inf., Brasília v. 38, n. 2, p. 56-68, 2009.

(RAWLINGS et al., 2002) RAWLINGS, A.; VAN ROSMALEN, P.; KOPER, R., RODRÍGUEZ-ARTACHO, M.; LEFRERE, P. **Survey of Educational Modelling Languages**. CEN/ISSS WS/LT Learning Technologies, 2002.

(REATEGUI, CAZELLA, 2005) REATEGUI E.; SILVIO C. **Sistemas de Recomendação**. XXV Congresso da Sociedade Brasileira de Computação, p. 306 - 348, 2005.

(REICHLING; VEITH; WOLF, 2007) REICHLING, T.; VEITH, M.; WULF, V. **Expert Recommender: Designing for a Network Organization**. Journal of Computer Supported Cooperative Work (JCSCW), p 1 -34, 2007.

(REIGELUTH, 1979) REIGELUTH, C. M. **In search of a better way to organize instruction: The elaboration theory**. Journal of Instructional Development, v. 2, 8-15, 1979.

(REIGELUTH, 1999) REIGELUTH, C. **Instructional Design theories and Models. A new Paradigm of Instrucional Theory**, 1999.

(REIGELUTH; STEIN, 1983) REIGELUTH, C.; STEIN, F. **The elaboration theory of instruction**. In C. Reigeluth (ed.), *Instructional Design Theories and Models*. Hillsdale, NJ: Erlbaum Associates, 1983.

(RESNICK; VARIAN, 1997) **Recommender Systems**. *Communications of the ACM*, New York, v. 40, n. 3, pp. 55-58, Março, 1997.

(ROBILLARD; WALKER, 2014) ROBILLARD, M.; WALKER, R. **An Introduction to Recommendation Systems in Software Engineering**. Springer-Verlag Berlin Heidelberg, 2014)

(RODRÍGUEZ-ARTACHO et al., 1999) RODRÍGUEZ-ARTACHO, M., VERDEJO, F., MAYORGA, J., & CALERO, M. **Using a High-Level Language to Describe and Create Web-Based Learning Scenarios**. IN: *IEEE Frontiers in Education Conference (FIE)*, November 10-13, 1999, San Juan, Puerto Rico USA. *Anais...*, 1999.

(RSLP, 2015) RSLP. **A Stemming Algorithm for the Portuguese Language**. Disponível em: <http://www.inf.ufrgs.br/~viviane/rsip/>. Acesso em 2015

(SEMAG, 2015) SEMAG. **Knowledge Discovery in the Internet**. *The Magic of Semantic Search*, 2015. Disponível em: <http://www.semagix.com/web-and-semantic-search/knowledge-discovery-in-the-internet.htm>

(SENGE, 2010) SENGE, P. M. **A Quinta Disciplina: Arte e Prática da Organização que Aprende**. 26. ed. Rio de Janeiro: Bestseller, 2010, 530 p.

(SENGE et al., 1994) SENGE, P.; KLEINER, A.; ROBERTS C.; ROSS, R.; SMITH, B. J. **The Fifth Discipline Field Book**. New York: Doubleday, 1994.

(SHAHZAD, 2009) SHAHZAD, S. **Learning from Experience: The Analysis of an Extreme Programming Process**. IN: *INTERNATIONAL CONFERENCE*

ON INFORMATION TECHNOLOGY, NEW GENERATIONS, 6, 2009. **Anais...**, 2009, p. 1405-1410.

(SHAPIRO, 1986) SHAPIRO, M. **Structure and Encapsulation in Distributed Systems: the Proxy Principle**. IN: 6<sup>TH</sup> Int. Conf. on Distributed Computing Systems, p. 198-204, 1986.

(SICILIA; LYTRAS, 2005) SICILIA, M.; LYTRAS, M. **The semantic learning organization**. Learning Organization, v. 12, n. 5, p. 402-410, 2005.

(SLAVIN, 1995) SLAVIN, R. Cooperative learning: **Theory, research, and practice**. Boston: Allyn & Baco, 1995.

(SQUID, 2015) SQUID Cache. **Squid: Optimizing Web Delivery**. Disponível em: <http://www.squid-cache.org/>

(STATA, 1989) STATA, R. **Organizational Learning – The Key to Management Innovation**. Sloan Management Review, v. 30, p. 63-74, 1989.

(STOREY et al., 2006) STOREY, M. A.; CHENG, L. T.; BULL, I.; RIGBY, P. **Shared waypoints and social tagging to support collaboration in software development**. IN: ANNIVERSARY CONFERENCE ON COMPUTER SUPPORTED COOPERATIVE WORK, NEW YORK, NEW YORK, USA: ACM, 2006. Anais...2006, p.195–198.

(TACLA; BARTHÈS, 2003) TACLA, Cesar A.; BARTHÈS, Jean-Paul, **A Multi-agent System For Acquiring and Sharing Lessons Learned**, Computers in Industry, v. 52, pp. 5 – 16, 2003.

(TANEV; MAGNINI, 2008) TANEV, H., MAGNINI, B. **Weakly Supervised Approaches for Ontology Population**. Proceeding of the conference on Ontology Learning and Population: Bridging the Gap between Text and Knowledge. Amsterdam, Netherlands, 2008.

(TAKEISHI, 2002) TAKEISHI, A. **Knowledge Partitioning in the Interfirm Division of Labor: The Case of Automotive Product Development.** Organization Science, v. 13, n. 3, p. 321-338, 2002.

(TEECE; PISANO; SHUEN, 1997) TEECE, D. J.; PISANO, G.; SHUEN, A. **Dynamic Capabilities and Strategic Management.** Strategic Management Journal, v.18, n.7, p. 509-533, 1997.

(TIRONI et al., 2015) TIRONI, H. U. M.; MENOLLI, A.; REINEHR, S.; MALUCELLI, A. **An approach to identify relevant subjects for supporting the learning scheme creation task.** In: Software Engineering and Knowledge Engineering International Conference (SEKE), 2015.

(TOGNERI et al., 2003) TOGNERI, D. F.; BRITO, S.; FALBO, R.; TAVARES, O.; MENEZES, C. **Um ambiente para aprendizagem cooperativa de engenharia de requisitos orientado a projetos.** IN: INTERNATIONAL CONFERENCE ON ENGINEERING AND COMPUTER EDUCATION (ICECE2003), SÃO PAULO, 2003. Anais... 2003, p. 1-5.

(TRAVASSOS; GUROV; AMARAL, 2002) TRAVASSOS, G. H.; GUROV, D.; AMARAL, E. G. A. **Introdução à Engenharia de Software Experimental.** Programa de Engenharia de Sistemas e Computação COPPE/UFRJ, 2002.

(TRINDADE, 2006) TRINDADE, A. L. P. **Uma Contribuição de para o Entendimento do Papel da Ensinagem na Preservação do Conhecimento em Ambientes de Fábrica de Software.** 2006, Tese (Doutorado) - Escola Politécnica, Universidade de São Paulo (USP), São Paulo, 2006.

(TSENG; HUANG, 2011) TSENG, S.-M.; HUANG, J.-S. **The correlation between Wikipedia and knowledge sharing on job performance.** Expert Systems with Applications, v. 38, n.5, p. 6118–6124. 2011.

(VALASKI; MALUCELLI; REINEHR; SANTOS, 2011) VALASKI, J.; MALUCELLI, A.; REINEHR, S.; SANTOS, R. **Ontology do Classify Learning**

**Material in Software Engineering Knowledge Domain.** ONTOBRAS-MOST, v. 776, pp. 37 – 47, 2011.

(VALENCIA; SIM, 2009) VALENCIA, R.; SIM, S. **Internet-scale code search.** In **Search-Driven Development-Users, Infrastructure, Tools and Evaluation**, p. 49 - 52, 2009.

(VASENSKA, 2013) VAZENSKA, Ivanka. **Organizational Learning and Employee Empowering Increasing Tourist Destination Performance.** IN: Management, Knowledge and Learning International Conference, 2013.

(W3C, 2008) W3C. **SPARQL Query Language for RDF.** 2008. Disponível em: <<http://www.w3.org/TR/rdf-sparql-query>>. Acesso em: 06 Abr. 2014.

(W3 SCHOOLS, 2015a) W3C Schools. **ASP QueryString Collection.** Disponível em: [http://www.w3schools.com/asp/coll\\_querystring.asp](http://www.w3schools.com/asp/coll_querystring.asp). Acesso em: 13/07/2015.

(W3 SCHOOLS, 2015b) W3C Schools. **Introduction to XML.** Disponível em: [http://www.w3schools.com/xml/xml\\_whatism.asp](http://www.w3schools.com/xml/xml_whatism.asp). Acesso em: 13/07/2015.

(WILSON; COLE, 1992) WILSON, B.; & COLE, P. **A critical review of elaboration theory.** **Educational Technology Research and Development**, v. 40, p. 63-79, 1992.

(WIRTH, 2000) WIRTH, R. **CRISP-DM: Towards a standard process model for data mining.** IN: Proceedings of Fourth International Conference on the Practical Application of Knowledge Discovery and Data Mining, 2000.

(WITTEN; MOFFAT; BELL, 1999) WITTEN, I. H.; MOFFAT, A.; BELL, T. C. **Compressing and Indexing Documents and Images.** Morgan Kaufmann Publishers, 1999.

(ZAIANE et al., 1998) ZAIANE, O. R.; XIN, M.; HAN, J. **Discovering Web Access Patterns and Trends by Applying OLAP and Data Mining Technology on Web Logs**. Proc. Advances in Digital Libraries Conf., pp. 19-29, 1998.

(ZELKOWITZ; WALLACE; BINKLEY, 2003) ZELKOWITZ, M.V.; WALLACE, D.R.; BINKLEY, D. W. **Experimental validation of new software technology**. IN: LECTURE NOTES ON EMPIRICAL SOFTWARE ENGINEERING, SOFTWARE ENGINEERING AND KNOWLEDGE ENGINEERING, RIVER EDGE, USA, 12, 2003. Anais..., 2003, p. 229-263.

(ZHAO; CHENG; ZHANG, 2010) ZHAO, C.; CHENG, M.; ZHANG, W. W. **Construction of Informal Learning Environment Model Based on SNS**. INTERNATIONAL SYMPOSIUM ON KNOWLEDGE ACQUISITION AND MODELING. Anais..., 2010, p. 247-249.

(ZOUAQ; NKAMBOU , 2009) ZOUAQ, A.; NKAMBOU, R. **Enhancing Learning Objects with an Ontology-Based Memory**. IEEE Transactions on Knowledge and Data Engineering, v. 21, n.6, p. 881-893, 2009.

## APÊNDICE A – QUESTIONÁRIO APLICADO NO EXPERIMENTO DE AVALIAÇÃO DA ABORDAGEM DE IDENTIFICAÇÃO DE TEMAS E PAPÉIS

Questionário aplicado aos funcionários que contribuíram com as pesquisas efetuadas na primeira fase do experimento de coleta de pesquisas.

Respondente (Nome):

Este questionário é sobre os resultados obtidos a partir dos exemplos de pesquisa fornecidos por você e que depois foram agrupados no sistema com as pesquisas de outros funcionários.

Por favor, responda a todas as questões. Se você não tem certeza sobre que resposta dar em uma questão, por favor, escolha entre as alternativas a que lhe parece mais apropriada. Esta, muitas vezes, poderá ser sua primeira escolha.

Procure responder a esta pesquisa de forma individual, conscienciosa e independente. A veracidade das suas respostas é fundamental. Em cada questão, marque apenas uma resposta, ou seja, aquela que melhor corresponda às suas escolhas. Os dados serão tratados estatisticamente, de modo a garantir absoluto sigilo a respeito das informações individuais prestadas.

1. Você utiliza sites de busca como Google, Yahoo e Bing para encontrar resoluções para as dificuldades no dia-a-dia?

Não       Muito Pouco       Indiferente       Muito       Completamente

2. A empresa contribui com a sua pesquisa diária deixando liberado na rede os acessos à fóruns, blogs ou wikis?

Não       Muito Pouco       Indiferente       Muito       Completamente

3. Você no dia-a-dia procura respostas para dúvidas pontuais (pequenas dúvidas que impedem o andamento da tarefa) primeiramente com os colegas de trabalho antes de fazer a pesquisa na internet?

Não       Muito Pouco       Indiferente       Muito       Completamente

4. O seguinte busca foi realizada por um ou mais respondentes: "C# MVC controller action result json result". Os temas envolvidos nesta busca tem ligação com as tecnologias do seu dia-a-dia? (Assunto mais pesquisado)

Não       Muito Pouco       Indiferente       Muito       Completamente

5. Você considera que é importante a empresa ter conhecimento sobre quais as dúvidas o funcionário tem durante a realização de uma tarefa?

Não       Muito Pouco       Indiferente       Muito       Completamente

6. Você evitaria fazer as buscas de dúvidas na internet se soubesse que elas estão sendo coletadas para melhor entendimento das dificuldades e criação de cursos?

Não       Muito Pouco       Indiferente       Muito       Completamente

7. Você acredita que executaria de maneira mais eficaz seu trabalho se houvessem cursos com temas voltados para os projetos internos?

Não       Muito Pouco       Indiferente       Muito       Completamente

8. Neste experimento cerca de 200 pesquisas diferentes foram registradas. Dentre as pesquisas o seguinte resultado foi encontrado: "Index Online SQL Server". Você considera este assunto importante no seu dia-a-dia? (Assunto pesquisado apenas uma vez)

Não       Muito Pouco       Indiferente       Muito       Completamente

9. Dentre os temas sugeridos para criação de cursos, foi sugerido "C# json request behaviour to allowget MVC3". Você concorda que há relação entre todas as palavras do tema?

Não       Muito Pouco       Indiferente       Muito       Completamente

10. Você se considera apto para lecionar um curso dentro da empresa com o seguinte tema: "C# SQL data reader WCF Service Adapter"?

Não       Muito Pouco       Indiferente       Muito       Completamente



## APÊNDICE B – PROJETO DE SISTEMA FICTÍCIO – PROCEDIMENTO DE EXPERIMENTO APLICADO NO EXPERIMENTO DE AVALIAÇÃO DA ABORDAGEM DE CRIAÇÃO DE CURSOS

Experimento para Avaliação – Dissertação de Mestrado em Engenharia de Software

Aluno: Huander Tironi

Orientadora: Profa. Doutora Andreia Malucelli

Co-orientador: Prof. Doutor André Menolli

Neste experimento será apresentado um projeto fictício de desenvolvimento de um sistema ERP. Após o entendimento da proposta de projeto de desenvolvimento, o participante deverá levantar as dúvidas relacionadas à:

- Regras de negócio (Cálculos, limitações do usuário, etc.);
- Padrões de Programação;
- Arquitetura do sistema;
- Tecnologias exigidas no desenvolvimento; e
- Outros;

Posteriormente o participante deverá acessar o Google (google.com.br) e buscar as respostas para todas as questões levantadas. O objetivo do experimento é registrar TODAS as pesquisas que foram realizadas no Google que foram necessárias para sanar as dúvidas levantadas. Abaixo um exemplo:

Dúvidas levantadas na leitura do projeto de desenvolvimento:

- *O que é linha de tendência?*
- *Como Instanciar objetos em javascript?*
- *Como desenvolver um projeto em WebAPI.*

Pesquisas no Google (Todas as pesquisas realizadas com o objetivo de entender as dúvidas):  
*“Linha de tendência”; “Exemplo de linha de tendência”; “Cálculo linha de tendência”; “linha de tendência gráfico”; “Instanciar objeto em javascript”; “JS criar objeto e instanciar”; “Exemplo de projeto WebAPI”; “WebAPI Java”; “WebAPI Cross Domain”;*

O objetivo deste experimento é coletar as buscas de todos os participantes para poder traçar um perfil de treinamento referente ao projeto apresentado.

**PROJETO DE DESENVOLVIMENTO DE SISTEMA ERP**

Uma **empresa do ramo de engenharia** solicita a uma fábrica de software na qual você é colaborador, um sistema ERP para que possam realizar atendimentos e vistorias em torres de telecomunicações.

O sistema conterà três (3) módulos principais: (i) módulo de mapeamento de torres; (ii) módulo de enlaces entre torres de telecomunicação; e (iii) módulo de manutenção de torres;

#### REQUISITOS DE NEGÓCIO

Módulo (i): O mapeamento ocorre por de um Tablet. O engenheiro vai até a torre de telecomunicação, e deverá utilizar o módulo para cadastrar a torre. Somente após cadastrada a torre o engenheiro poderá cadastrar antenas. O sistema deverá estar preparado para receber qualquer tipo de antena. Para cada antena o engenheiro poderá calcular o “Laudo Radiométrio”. O sistema deverá realizar o cálculo e gerar um arquivo PDF com o memorial de cálculo para poder ser enviado por e-mail para o cliente.

Módulo (ii): O módulo de enlaces será realizado por computador. O engenheiro terá um mapa onde poderá selecionar a latitude e longitude de diferentes torres. O sistema deverá calcular o Azimute de cada antena para que elas possam se comunicar.

Módulo (iii): O módulo de manutenção será acessado pelo cliente. O cliente entrará a informar que há algum problema em sua estação. Ao realizar a solicitação o engenheiro recebe o aviso para fazer uma visita. O cliente poderá acessar o módulo de manutenção por celular (Apache Cordova) .

#### REQUISITOS TÉCNICOS

As aplicações de tablet, celular e desktop deverão ser WebAPPs desenvolvidos com Javascript (KnockoutJS e JQuery) em SPA. Utilizar o Framework Apache Cordova para Tablet e Celular.

O Sistema deverá possuir uma WebAPI em um servidor em nuvem. Esta WebAPI deverá ser desenvolvida em C# MVC na VERSÃO MAIS ATUALIZADA.

A comunicação entre os Aplicativos e a WebAPI deverá ser via AJAX.

O banco de dados será o MS SQL Server. A comunicação dos dados de listas para o banco será via JSON. O DER deverá ser documentado.

**Coloque aqui as suas dúvidas:**

**Coloque aqui as buscas realizadas no Google:**