

Universidade Federal do Piauí
Campus Senador Helvídio Nunes de Barros
Curso Bacharelado em Sistemas de Informação

José Klisanderson de Sousa

Um Mapeamento Sistemático das Técnicas, Métodos e Ferramentas Utilizadas em Aprendizado de Ontologias

PICOS
2013

José Klisanderson de Sousa

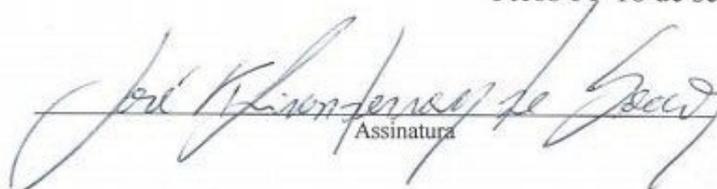
Um Mapeamento Sistemático das Técnicas, Métodos e Ferramentas Utilizadas em
Aprendizado de Ontologias

Trabalho de Conclusão de Curso apresentado ao Curso de Bacharelado em Sistemas de Informação do Campus Senador Helvídio Nunes de Barros da Universidade Federal do Piauí como parte dos requisitos para obtenção do Grau de Bacharel em Sistemas de Informação, sob orientação da Professor Esp. Dennis Sávio M. da Silva

PICOS
2013

Eu, **José Klisanderson de Sousa**, abaixo identificado(a) como autor(a), autorizo a biblioteca da Universidade Federal do Piauí a divulgar, gratuitamente, sem ressarcimento de direitos autorais, o texto integral da publicação abaixo discriminada, de minha autoria, em seu site, em formato PDF, para fins de leitura e/ou impressão, a partir da data de hoje.

Picos-PI 18 de setembro de 2013.


Assinatura

FICHA CATALOGRÁFICA

Serviço de Processamento Técnico da Universidade Federal do Piauí
Biblioteca José Albano de Macêdo

S725m Sousa, José Klisanderson de.
Um Mapeamento sistemático das técnicas, métodos e ferramentas utilizadas em aprendizado de ontologias / José Klisanderson de Sousa. – 2013.
CD-ROM : il. ; 4 ¾ pol. (70 p.)

Monografia(Bacharelado em Sistemas de Informação) –
Universidade Federal do Piauí. Picos-PI, 2013.
Orientador(A): Prof. Esp. Dennis Sávio Martins da Silva

1. Aprendizado de Ontologias. 2. Revisões Sistemáticas.
3. Mapeamento Sistemático de Literatura. I. Título.

CDD 004

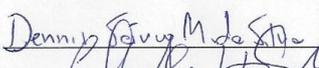
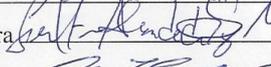
José Klisanderson de Sousa

Um Mapeamento Sistemático das Técnicas, Métodos e Ferramentas Utilizadas em
Aprendizado de Ontologias

Trabalho de Conclusão de Curso apresentado ao Curso de Bacharelado em Sistemas de Informação do Campus Senador Helvídio Nunes de Barros da Universidade Federal do Piauí como parte dos requisitos para obtenção do Grau de Bacharel em Sistemas de Informação, sob orientação da Professor Esp. Dennis Sávio M. da Silva

Data de Aprovação:

16/09/13

Dennis Sávio Martins da Silva  UFPI-CSHNB
Ivenilton Alexandre de Souza Moura  UFPI-CSHNB
Ismael de Holanda Leal  UFPI-CSHNB

PICOS

2013

A DEUS,
à Alaide, minha querida mãe pelo exemplo de vida. Pessoa ao qual
sempre me apoiou, incentivou e acreditou nas minhas decisões de vida.
Você é meu porto seguro.

Agradeço primeiramente a DEUS pelo suporte espiritual e por manter a minha sanidade mental me proporcionando força e determinação para realização desse trabalho e das minhas demais conquistas.

À minha família pelo apoio, dedicação e estímulo que dispuseram para à minha formação. Em especial, a minha mãe Alaide Joaquina de Jesus, meu padrasto Antônio Carlos dos Santos Alencar, meu irmão Francisco Kairo de Jesus Alencar, meu Tio João José de Sousa e meus avôs José Manoel de Sousa e Joaquina Vitória de Jesus.

À meu orientador, Dennis Sávio Martins da Silva, pela amizade, pela orientação, pelo incentivo e pela disposição em dividir seu tempo e conhecimento.

Obrigado aos membros da banca examinadora, Dennis Sávio Martins da Silva, Ivenilton Alexandre de Souza Moura e Ismael de Holanda Leal, pelas sugestões e discussões durante a defesa, que enriqueceram muito este trabalho.

À todos os demais professores, que contribuíram direto ou indiretamente para a minha formação, desde o jardim de infância até a conclusão desse curso superior.

À meu grande amigo Ryan Ribeiro de Azevedo, pela amizade, pela orientação, pelo incentivo e pelo interesse constante no progresso do meu trabalho. Sendo ele, um dos responsáveis pela minha evolução. Enfim, por propiciar o meu envolvimento no seu mundo de sabedoria.

Aos meus amigos "sem os quais eu nada seria" e que com seus sorrisos, palavras de incentivo e carinho tornaram minha caminhada mais suave – Jaciomar (*In memoriam*), Pâmela, Danila, Janaina,...

Obrigado a todos que, mesmo não tendo seus nomes citados aqui, contribuíram para a realização deste trabalho de alguma forma.

E obrigado a DEUS, sempre, pela minha vida.

“Que os vossos esforços desafiem as impossibilidades, lembrai-vos de que as grandes coisas do homem foram conquistadas do que parecia impossível.”

Charles Chaplin

“A menos que modifiquemos a nossa maneira de pensar, não seremos capazes de resolver os problemas causados pela forma como nos acostumamos a ver o mundo.”

Albert Einstein

“Feliz aquele que transfere o que sabe e aprende o que ensina.”

Cora Coralina

Resumo

Esta pesquisa selecionará os trabalhos mais relevantes da área de Aprendizado de Ontologias (AO) a partir do mecanismo de estudo empírico na forma de um Mapeamento Sistemático de Literatura (MSL) que é um dos tipos de Revisões Sistemáticas. O intuito deste trabalho é o de evidenciar a interseção das técnicas, métodos e ferramentas utilizados na construção automática de ontologias. Para isso serão utilizados os engenhos de busca de quatro bibliotecas digitais: *IEEE Computer Society Digital Library*, *ACM*, *Science Direct* e *Scopus* para buscas automatizadas, além de busca manual em *journals* relevantes para a área.

Palavras-chave: Aprendizado de Ontologias, Ferramentas, Mapeamento Sistemático de Literatura, Métodos, Revisões Sistemáticas, Técnicas.

Abstract

This research will select the more relevant works on Ontology Learning (OL), by a Systematic Literature Review, a empiric study mechanism, which is from the Systematic Review type. The target of this work is show the main techniques, methods and tools used on the automatic construction of ontologies. For this, will be used search engines for automatic search from four Digital Libraries: IEEE Computer Society Digital Library, ACM, Scopus and Science Direct, and manual search on journals relevant from the area.

Keywords: Ontology Learning, Tools, Systematic Literature Mapping, Methods, Systematic reviews, Techniques.

Lista de Figuras

Figura 1 -	Processos de condução de uma RS	19
Figura 2 -	Etapas da condução de um MSL	23
Figura 3 -	Detalhe das etapas de seleção dos documentos	23
Figura 4 -	Exemplo de uma ontologia aplicada a Engenharia do Conhecimento	27
Figura 5 -	Camadas de AO	29
Figura 6 -	Hierarquia de Conceitos	32
Figura 7 -	Etapas de Pesquisa	41
Figura 8 -	Resultados da busca automatizada.	48
Figura 9 -	Resultados da busca manual.	48
Figura 10 -	Resultados da busca automatizada e manual.	49
Figura 11 -	Resultado seleção de trabalhos primários que compõem o estudo.	49
Figura 12 -	Evolução do processo de seleção dos estudos.	50
Figura 13 -	Distribuição dos estudos ao longo dos anos.	50
Figura 14 -	Distribuição dos estudos pelos países das instituições dos pesquisadores.	51
Figura 15 -	Frequência de aparecimento das ferramentas nos estudos primários	56

Lista de Tabelas

Tabela 1 -	Revisão Sistemática versus Revisão Tradicional	20
Tabela 2 -	Vantagens e Desvantagens da Revisão Sistemática	20
Tabela 3 -	Termos e Sinônimos	24
Tabela 4 -	Técnicas de aprendizado de ontologias a partir de texto	35
Tabela 5 -	Tipos de abordagem metodológica	37
Tabela 6 -	Quadro Metodológico da Pesquisa	37
Tabela 7 -	Classificação segundo Cooper	39
Tabela 8 -	Descrição da cobertura do MS	40
Tabela 9 -	Organização dos trabalhos e resultados	40
Tabela 10 -	Termos e Sinônimos	44
Tabela 11 -	Fontes para busca automática	44
Tabela 12 -	Fontes para busca manual (<i>Journals</i>)	44
Tabela 13 -	Informações coletadas dos estudos primários	46
Tabela 14 -	Informações coletadas dos estudos selecionados	46
Tabela 15 -	Métodos para aprendizado de ontologias	52
Tabela 16 -	Técnicas para aprendizado de ontologias	53
Tabela 17 -	Ferramentas para aprendizado de ontologias	55
Tabela 18 -	Estudos primários retornados na busca automática	65
Tabela 19 -	Estudos primários retornados na busca manual	66

Lista de abreviaturas e siglas

ACE	Extração Automático de Conteúdo
AM	Aprendizagem de Máquinas
AO	Aprendizado de Ontologias
EMS	Estudo de Mapeamento Sistemático
EO	Engenharia de Ontologias
EP	Estudos Primários
MSL	Mapeamento Sistemático da Literatura
OL	Ontology Learning
PhD	Philosophiae Doctor
PLN	Processamento de Linguagem Natural
TFIDF	Term Frequency Inverse Document Frequency

Sumário

1	Introdução	15
1.1	Motivação	15
1.2	Objetivos	15
1.2.1	Objetivos Gerais	15
1.2.2	Objetivos Gerais	16
1.3	Formulação da Hipótese de Trabalho	16
1.4	Estrutura da Monografia	16
2	Revisão Sistemática, Revisão Sistemática da Literatura, Mapeamento Sistemático da Literatura	18
2.1	Revisão Sistemática	18
2.1.1	Processo de Condução de uma RS	18
2.1.2	Diferenças entre Revisão Tradicional e Revisão Sistemática	19
2.1.3	Vantagens e Desvantagens de uma Revisão Sistemática	20
2.2	Revisão Sistemática da Literatura	21
2.2.1	Objetivo da RSL	21
2.2.2	Pontos importantes para realizar uma RSL	21
2.2.3	Critérios para conduzir uma RSL	21
2.3	Mapeamento Sistemático da Literatura	22
2.3.1	Objetivo do MSL	22
2.3.2	Etapas de condução de um MSL	23
2.3.3	Considerações Finais	25

3	Aprendizado de Ontologias a partir de Texto: Camadas, Métodos, Técnicas e Ferramentas	26
3.1	Motivação	26
3.2	Ontologias	27
3.2.1	Características das ontologias	27
3.3	Aprendizado de Ontologias	28
3.3.1	Camadas de Aprendizado de Ontologias	29
3.3.2	Tarefas do Aprendizado de Ontologias	29
3.3.3	Métodos para Aprendizado de Ontologias a partir de Texto	33
3.3.4	Técnicas para Aprendizado de Ontologias	35
3.3.5	Ferramentas para Aprendizado de Ontologias	35
3.4	Considerações Finais	36
4	Metodologia	37
4.1	Classificação da Pesquisa	37
4.1.1	Classificação segundo Cooper	38
4.2	Ciclo da Pesquisa	40
4.2.1	Mapeamento Sistemático	41
4.2.2	Escopo e Questão de Pesquisa	42
4.2.3	Estratégia de Busca	43
4.2.4	Critérios de Inclusão e Exclusão e Procedimentos de Seleção dos estudos no MSL	44
4.3	Considerações Finais	46
5	Resultados	47
5.1	Extração e Análise dos Dados	47
5.2	Mapeamento das Evidências	51
5.2.1	Métodos	52

5.2.2	Técnicas	52
5.2.3	Ferramentas	54
5.3	Discussão dos Resultados	56
5.4	Considerações Finais	57
6	Considerações Finais	58
6.1	Ameaças a Validade	58
6.2	Trabalho Futuro	58
6.3	Conclusão	59
	Referências	60
	Apêndice A – Calibração das <i>Strings</i> de Buscas aplicadas na busca automatizada e resultados da busca manual	65
A.1	Busca Automática	65
A.2	Busca Manual	66
	Apêndice B – Estudos Primários	67

1 Introdução

1.1 Motivação

No contexto do desenvolvimento de ontologias, diversos métodos, processos e técnicas de desenvolvimento surgiram e vem emergindo ao longo da última década. Dentre estes artefatos de construção de ontologias, estão os métodos manuais e os métodos automáticos ou semiautomáticos. O objetivo deste trabalho é realizar um Mapeamento Sistemático da Literatura (MSL) em relação aos métodos automáticos de construção de ontologias. A essa área de pesquisa, dá-se o nome de Aprendizado de Ontologias (AO) do inglês *Ontology Learning* (OL). O intuito é o de evidenciar lacunas em relação ao estado da arte da área de AO, além de mapeá-las em obra única.

Os estudos empíricos, tal como o MSL, é uma ferramenta pertinente para evidenciar fatos e lacunas em aberto, o MSL provê um modo sistemático, disciplinado, computável e controlado para avaliação e validação das lacunas em aberto e mapeamento do estado da arte, evitando o discurso modesto de intuição, opinião e especulação dado de forma pessoal ou não sistêmico por pesquisadores. Assim, evita-se o conhecimento não confiável. A importância do MSL vem da necessidade de evidências sobre eficiência e eficácia dos métodos automáticos de construção de ontologias, uma vez que, os mesmos estão sendo usados e pesquisados por diversos grupos de pesquisa ao redor do planeta.

Assim, é possível obter uma base de conhecimento confiável que possa ser utilizada para apoiar a tomada de decisões, que deve ser usada no decorrer do desenvolvimento de soluções em engenharia de ontologias quando se fizer necessário.

1.2 Objetivos

1.2.1 Objetivos Gerais

O objetivo desse trabalho é realizar um mapeamento sistemático da literatura a fim de evidenciar as técnicas, métodos e ferramentas utilizados no aprendizado de ontologias.

1.2.2 Objetivos Gerais

Como resultado proveniente do mapeamento sistemático realizado, pretende-se alcançar os seguintes objetivos específicos:

- Indicar as técnicas e métodos de aprendizado de ontologias existentes.
- Indicar as ferramentas existentes que utilizam técnicas e métodos de aprendizado de ontologias.

1.3 Formulação da Hipótese de Trabalho

Não teremos hipótese de trabalho, trata-se de uma pesquisa qualitativa que responderá as seguintes Questões de Pesquisa (QP):

- **QP1: Quais são os métodos utilizados no Aprendizado de Ontologias?**

Nesse ponto, o mapeamento sistemático (MS) pretende responder quais os métodos existentes para o aprendizado de ontologias.

- **QP2: Quais são as técnicas utilizadas no Aprendizado de Ontologias?**

Nesse ponto, o mapeamento sistemático (MS) pretende responder quais as técnicas existentes para o aprendizado de ontologias.

- **QP3: Quais são as ferramentas utilizadas no Aprendizado de Ontologias?**

Nesse ponto, o mapeamento sistemático (MS) pretende responder quais as ferramentas existentes para o aprendizado de ontologias.

1.4 Estrutura da Monografia

Após a introdução, que relatou sobre as motivações e objetivos para o desenvolvimento do presente trabalho serão apresentados os próximos capítulos, que estão organizados da seguinte forma:

- Capítulo 2 - É apresentada a primeira parte do referencial teórico, contendo os conceitos sobre Revisão Sistemática, Revisão Sistemática da Literatura e Mapeamento Sistemático da Literatura.
- Capítulo 3 - É apresentada a parte complementar do referencial teórico, com conceitos no qual o presente trabalho está fundamentado. Onde se descreve sobre Aprendizado de Ontologias a partir de texto, apontando as camadas, métodos, técnicas e ferramentas utilizadas na construção de ontologias.

- Capítulo 4 - É descrita a metodologia que foi usada como embasamento para realizar esta pesquisa. É apresentada a classificação da pesquisa contendo a apresentação de um quadro metodológico, as etapas principais deste estudo, os métodos seguidos pelo mapeamento sistemático, apresentação do protocolo que foi definido, e como os dados foram extraídos e analisados.
- Capítulo 5 - É apresentado os resultados do mapeamento sistemático. No início são apresentadas informações gerais sobre o processo de pesquisa e seleção dos estudos primários que participaram do estudo, como principais fontes, número de estudos retornados, distribuição por países e temporal, e outras informações. Ao final do Capítulo são apresentadas as evidências encontradas que serviram para responder as perguntas de pesquisa e o mapeamento resultante.
- Capítulo 6 – Por fim, o capítulo de considerações finais mostrará as ameaças a validade do estudo, os objetivos que foram alcançados após o desenvolvimento do trabalho, com sugestões para trabalhos futuros.

2 Revisão Sistemática, Revisão Sistemática da Literatura, Mapeamento Sistemático da Literatura

Neste capítulo será apresentada a primeira parte da fundamentação teórica obtida através de uma revisão bibliográfica (informal). Os tópicos que serão abordados são: Revisão Sistemática (RS), Revisão Sistemática da Literatura (RSL) e Mapeamento Sistemático da Literatura (MSL). As subseções a seguir mostram em detalhes cada tópico citado.

2.1 Revisão Sistemática

A Revisão Sistemática trata-se de estudo secundário no qual o processo de pesquisa segue um conjunto de etapas metodologicamente bem definidas de acordo com um protocolo antecipado, onde sua adoção visa reduzir as diferenças relacionadas a uma revisão informal (KITCHENHAM et al., 2004) (BIOLCHINI et al., 2005). Por estudos secundários, entende-se a condução de estudos que visem a identificar, avaliar e interpretar todos os resultados relevantes a um determinado tópico de pesquisa, fenômeno de interesse ou questão de pesquisa (KITCHENHAM, 2004).

Biolchini et al. (2005), apresenta em seu relatório técnico sobre como conduzir revisões sistemáticas em Engenharia de Software (ES), a seguinte definição de Revisão Sistemática:

“Uma RS é usada para se referir a uma metodologia (sic) específica de pesquisa, desenvolvida de forma a encontrar e avaliar as evidências disponíveis pertencentes a um tópico focado” (BIOLCHINI et al., 2005, pag.1).

2.1.1 Processo de Condução de uma RS

As etapas do processo de revisão sistemática, compostas pelo planejamento da revisão, execução da revisão, análise e divulgação dos resultados descritos na Figura 1, não seguem necessariamente uma sequência de execução, uma vez em que estas passam por iterações, onde muitas das atividades são iniciadas na fase de planejamento e posteriormente serão refinadas. As fases de condução da revisão sistemática utilizadas neste trabalho foram adaptadas de tra-

balhos de referência citados na literatura (H.LITTELL et al., 2008) (BIOLCHINI et al., 2005) (KITCHENHAM et al., 2004) (PAI et al., 2004).

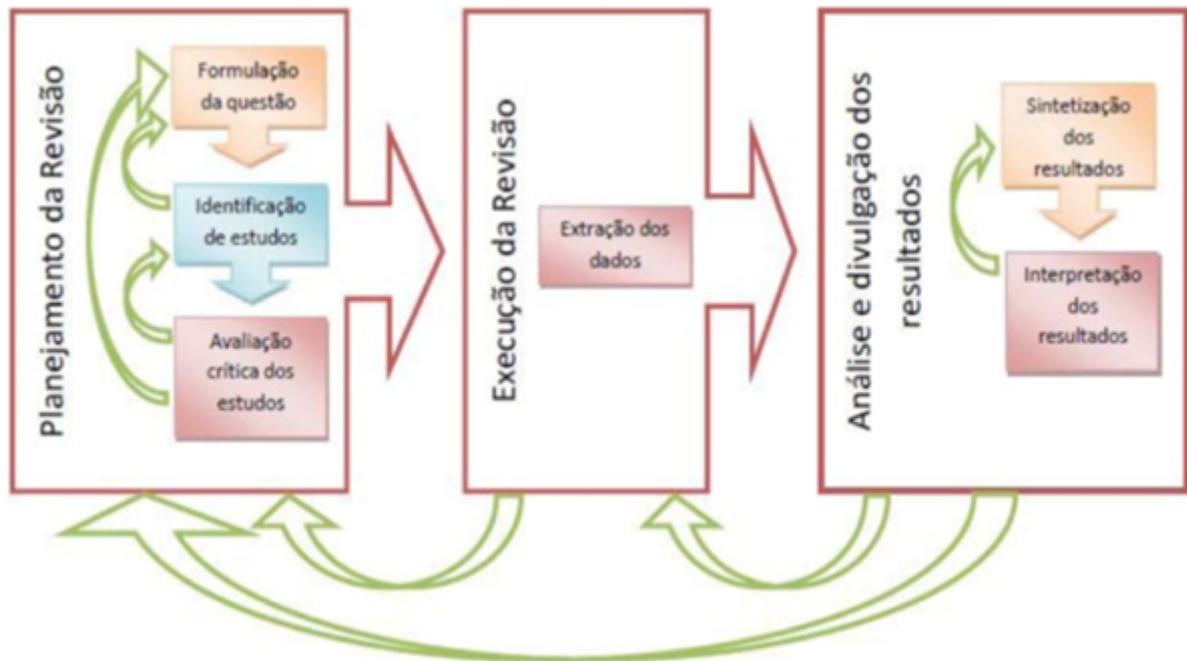


Figura 1 – Processos de condução de uma RS

Fonte: (BIOLCHINI et al, 2005; KITCHENHAM et al, 2004; PAI et al, 2004; LITTEL et al, 2008).

Na fase do Planejamento da Revisão, que engloba a formulação da questão de pesquisa, identificação dos estudos e conseqüentemente uma avaliação crítica dos estudos que foram identificados, é definido o objetivo da pesquisa assim como o protocolo para efetuar a revisão. Na fase de condução da revisão, é realizada a extração dos dados de acordo com os estudos que foram selecionados. É nessa etapa em que se identificam os estudos primários, que serão selecionados e em seguida avaliados. Na fase de análise e divulgação dos resultados que é composta pela síntese e interpretação dos resultados, ocorre o processo de extração e síntese dos dados dos estudos para que sejam finalmente publicados.

2.1.2 Diferenças entre Revisão Tradicional e Revisão Sistemática

Uma revisão tradicional se diferencia de uma revisão sistemática pelo fato desta utilizar para coletar e interpretar seus dados métodos informais e subjetivos. Vale ressaltar que esse método de revisão não descreve a busca, seleção e a avaliação da qualidade dos estudos. Por outro lado, em uma revisão sistemática é definido uma questão de pesquisa a ser analisada, e logo em seguida é realizado uma busca abrangente e exaustiva por estudos primários, que serão avaliados e sintetizados de acordo com um método explicitamente pré-definido (PAI et al., 2004). Para uma melhor compreensão observe a Tabela 1 onde são apresentadas algumas diferenças entre Revisão Sistemática e Revisão Tradicional:

Características	Revisão Tradicional	Revisão Sistemática
Questão	Escopo abrangente. formulação genérica.	Foco definido, formulação específica.
Identificação da pesquisa	Usualmente não especificada e tendenciosa.	Fontes abrangentes, estratégia de busca definida e explícita
Seleção	Usualmente não especificada e tendenciosa	Baseada em critérios explícitos e uniformemente aplicados.
Análise	Variável	Rigorosa nos objetivos e no método, e crítica.
Síntese	Geralmente um sumário qualitativo	Síntese qualitativa e/ou quantitativa
Inferências	Algumas vezes baseadas em evidência.	Sempre baseadas em evidência.

Tabela 1 – Revisão Sistemática versus Revisão Tradicional

2.1.3 Vantagens e Desvantagens de uma Revisão Sistemática

Esse tópico apresenta algumas das vantagens e desvantagens relacionadas à Revisão Sistemática, com o objetivo de repassar para o leitor as possíveis vantagens de sua utilização, assim como as dificuldades que poderá se deparar quando estiver efetuando esse tipo de revisão. Para que o leitor possa ter uma melhor compreensão sobre o assunto, ver a Tabela 2 Revisão Sistemática: Vantagens versus Desvantagens.

Vantagens	Desvantagens
Quanto menor o intervalo de confiança, maior a precisão dos resultados.	Não surte efeito direto nos estudos que compõem a RS, apenas faz recomendações para que estudos futuros não cometam erros.
Previne a duplicação de esforço, uma vez completada a RS ela não precisa ser repetida.	Consome muito tempo como em qualquer pesquisa científica de qualidade
Sua utilização evita controvérsia na literatura, dando importância à qualidade ao invés da quantidade de estudos realizados	É quase impossível realizar uma RS sozinho, pois é preciso de dois profissionais para avaliar os estudos.
Pode ser rapidamente atualizada, com inclusão de novos estudos publicados.	Envolve um grande trabalho intelectual, em toda a sua execução.
Utiliza uma metodologia reprodutível, portanto, científica.	Demanda esforço e dedicação.

Tabela 2 – Vantagens e Desvantagens da Revisão Sistemática

2.2 Revisão Sistemática da Literatura

É um tipo de RS, considerada um estudo secundário que representa uma abordagem diferente para as questões relevantes em uma área de pesquisa, possibilitando o surgimento de novas possibilidades e assim gerando novos tipos de conhecimento. Um dos primeiros passos para realização de uma RSL é definir um protocolo de pesquisa contendo todo o planejamento da revisão (CHARTERS; KITCHENHAM, 2007). A autora define RSL como uma forma de identificar, avaliar e interpretar todas as pesquisas disponíveis que são relevantes para uma particular questão de pesquisa, ou área, ou fenômeno de interesse.

2.2.1 Objetivo da RSL

O objetivo de uma revisão sistemática de literatura é apresentar uma avaliação justa do tópico de pesquisa, à medida que utiliza uma metodologia de revisão rigorosa, confiável e passível de auditoria (KITCHENHAM et al., 2004).

2.2.2 Pontos importantes para realizar uma RSL

Uma RSL deve ser iniciada com o desenvolvimento de um protocolo. No protocolo são especificados os pontos que o compõem, como descritos a seguir:

- Definir a pergunta científica onde será especificada a população de interesse;
- Identificar as bases de dados a serem analisadas, assim como uma estratégia de busca;
- Estabelecer critérios para selecionar os artigos encontrados;
- Aplicar os critérios de inclusão e exclusão de artigos encontrados;
- Analisar criticamente e avaliar todos os estudos incluídos na revisão;
- Preparar um resumo crítico sobre as informações disponibilizadas pelos artigos que foram incluídos na revisão;
- Apresentar uma conclusão informando evidências sobre o efeito da intervenção;

2.2.3 Critérios para conduzir uma RSL

O processo de condução de uma RSL é composto por cinco etapas (delimitação do problema de pesquisa, busca por investigações com o objetivo de evitar abordagens infrutíferas,

busca de *insights* metodológicos, a identificação de recomendações para pesquisas futuras e a busca por fundamentação teórica), de acordo com o campo de estudo essas etapas podem sobrepor-se ou ocorrer numa ordem diferente.

Randolph (2009) propõe os seguintes critérios para condução de uma Revisão Literária:

- **Delimitar o problema de pesquisa:** Nesta etapa são descobertas as questões de investigação que são viáveis para o estudo;
- **Buscar novas linhas de investigação, evitando abordagens infrutíferas:** Envolve a elaboração de uma proposta de pesquisa, contendo uma introdução sobre a revisão da literatura, um projeto de pesquisa, e o método de pesquisa a ser utilizado.
- **Busca de insights metodológicos:** Essa etapa tem por objetivo o desenvolvimento e testes dos métodos de coleta de dados.
- **Identificar recomendações pra pesquisas futuras:** Nessa fase realiza-se um estudo principal com o objetivo de identificar lacunas para pesquisas futuras.
- **Busca de fundamentação teórica:** Nessa fase é realizada uma visão crítica sobre os estudos que são significantes para o trabalho em que está sendo desenvolvido.

2.3 Mapeamento Sistemático da Literatura

O MSL é um tipo de revisão que complementa a Revisão Sistemática, (também chamado de escopo do estudo), é responsável por fornecer uma visão ampla sobre uma área que se deseja investigar, com o intuito de definir se existe evidência de pesquisa sobre um tema, e indicar a quantidade de evidências encontradas no tema que foi investigado. Os resultados adquiridos com a investigação indicam as áreas mais adequadas para se realizar revisões sistemáticas de literatura. Petersen et al. (2008) afirma que um mapa sistemático é um método definido de construir um esquema de classificação e estrutura em um campo de interesse.

2.3.1 Objetivo do MSL

Em um MSL devem estar definidos: as questões de pesquisa, a estratégia ou string de busca e as fontes de buscas, assim como os respectivos critérios de inclusão e exclusão de conteúdo do MSL. Petersen et al. (2008) ressalta que o objetivo principal desse tipo de estudo é:

“Prover uma visão geral de uma área de pesquisa, identificando a quantidade e o tipo de pesquisas e resultados disponíveis dentro dela. Geralmente, o pesquisador quer conhecer também as frequências de publicações ao longo do tempo para identificar tendências” (PETERSEN et al., 2008, pag.2).

2.3.2 Etapas de condução de um MSL

A condução de um MSL é realizada de acordo com o seguinte conjunto de etapas: definição da questão de pesquisa, o esclarecimento do objetivo da revisão, a busca por documentos relevantes a pesquisa, que serão selecionados de acordo com a leitura de seus respectivos resumos, para que sejam extraídos os dados, e por fim a realização do mapeamento. Cada etapa do processo tem um resultado, o resultado final do processo será o mapa sistemático. O processo é detalhado na Figura 2, a seguir:

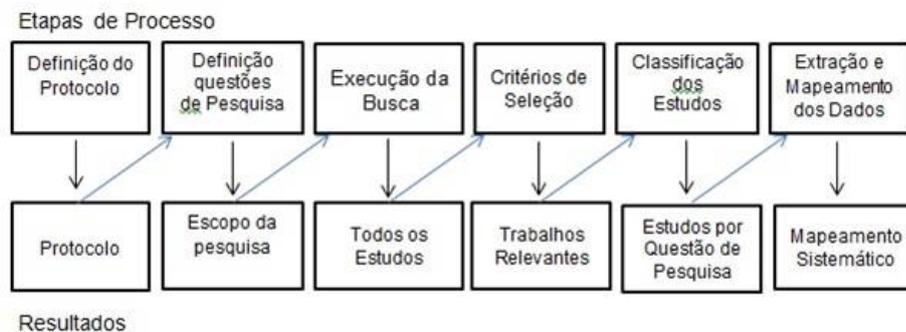


Figura 2 – Etapas da condução de um MSL

Fonte: Adaptado de (Petersen et al, 2008:2).

A Figura 3 abaixo detalha a etapa que realiza os critérios de seleção dos documentos:

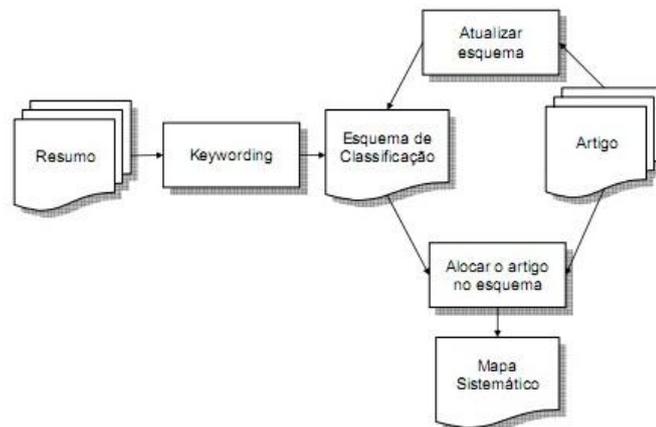


Figura 3 – Detalhe das etapas de seleção dos documentos

Fonte: Adaptado de (Petersen et al, 2008:4).

Os critérios de seleção dos documentos são compostos por duas etapas: A análise de artigos selecionados, e a definição de um esquema de classificação dos trabalhos, após essa fase eles serão separados em grupos de acordo com a questão que se deseja responder. Na segunda etapa, a lista de documentos selecionados, e já classificados podem sofrer alterações ou atualização, o que resulta na inserção ou exclusão de documentos em todo o decorrer da construção do MSL.

Estratégia de Busca

• *String* de Busca

A principal parte de estratégia de busca é a elaboração da *string* de busca, que é construída com base em termos selecionados da questão de pesquisa. A construção da *string* de busca seguir uma estrutura baseada em (KITCHENHAM et al., 2006), que consiste nos seguintes passos:

- Selecionar a partir da questão de pesquisa os termos para construir a *string*;
- Realizar a tradução desses termos para o inglês, por ser o idioma mais utilizado na literatura da Ciência da Computação;
- Identificar os sinônimos referentes aos termos selecionados na questão de pesquisa;
- A *string* de busca é gerada a partir da combinação desses termos principais e seus sinônimos;
- Utilizar o conectivo booleano *OR* para interligar os sinônimos;
- Utilizar o conectivo booleano *AND* para interligar os termos.

Verificar a *string* de busca construída realizando testes nas bibliotecas selecionadas e comparando a relevância dos documentos obtidos:

Termos	Sinônimos
<i>ontology learning</i>	
<i>Method</i>	<i>methodology, process</i>
<i>Technique</i>	<i>techniques</i>
<i>Tool</i>	<i>framework, plug-in</i>

Tabela 3 – Termos e Sinônimos

Com base na estratégia descrita acima foi obtida a seguinte *string* de busca:

("ontology learning") AND (framework OR method OR methodology OR process OR technique OR thecniques) AND (tool OR plug-in)

• Fontes de Busca

O processo utilizado para buscar os estudos primários pode ser realizado com buscas automatizadas através de engenhos de busca das bibliotecas digitais, onde será aplicada a *string* de busca que foi descrita, e de forma manual em *journals* relevantes à área. A busca manual é importante para dar mais confiança ao pesquisador que conduz a revisão sistemática, ampliar a cobertura da revisão, captar estudos que ainda não foram indexados

pelos engenhos de busca das bibliotecas digitais, devido à publicação recente, e identificar estudos relevantes que não usam os termos da *string* utilizada na busca automatizada. No caso da busca manual, o que determinou o período de tempo de busca foi a disponibilidade do material na Internet. Foram utilizados os seguintes critérios para selecionar as fontes de buscas:

- A fonte deve ter disponibilidade de consulta a artigos através da web;
- A fonte deve possuir mecanismos de busca através de palavras-chaves;
- Fontes de buscas automáticas já utilizadas em outras revisões da Literatura sobre *Ontology Learning*;

• **Inclusão e Exclusão de estudos no MSL**

A seleção dos estudos primários referentes à pesquisa segue o procedimento onde é realizada uma análise dos estudos obtidos através das bibliotecas digitais por meio da leitura de título, resumo e conclusão de cada documento retornado após aplicar a *string* de busca. A inclusão de um trabalho no mapeamento se dá pela sua relevância em relação à questão de pesquisa. Os critérios de inclusão e exclusão definidos para esse estudo são do tipo “restrição”, essa classificação de critério é originalmente proposta por (RABISER et al., 2010).

2.3.3 Considerações Finais

Neste capítulo foi apresentada uma breve abordagem sobre a metodologia que será utilizada para guiar a pesquisa, com o intuito de demonstrar ao leitor que a condução de um mapeamento sistemático que utiliza um protocolo de busca é uma forma de identificar, analisar e interpretar todos os estudos disponíveis sobre um determinado tema. Tendo por objetivo principal a descoberta de evidências sobre a determinada área de estudo.

3 Aprendizado de Ontologias a partir de Texto: Camadas, Métodos, Técnicas e Ferramentas

Neste capítulo será apresentada a parte complementar da fundamentação teórica, obtida através de uma revisão bibliográfica (informal) na qual o presente trabalho está fundamentado. É demonstrado um estudo sobre o estado-da-arte em AO, apresentando as camadas, métodos, técnicas e ferramentas utilizados no processo de construção de ontologias através de processos automáticos ou semi-automáticos aplicados em recursos existentes num domínio específico. As subseções a seguir demonstram em detalhes os tópicos explanados acima.

3.1 Motivação

O termo Aprendizado de Ontologias (AO) do inglês (*Ontology Learning*) é historicamente ligado a Web Semântica que foi introduzida por Berners-Lee et al. (2001) e surgiu como um importante sub-campo da Engenharia de Ontologias (EO) com o intuito de automatizar o processo de construção de ontologias.

AO foi introduzido por Maedche e Staab (2001) sendo descrito como o processo de aquisição de um modelo de domínio a partir de dados. Esse termo foi cunhado pelos autores Cimiano et al. (2009) refere-se à geração automática ou semi-automática de ontologias a partir de varias fontes de dados, incluindo folksonomias¹ e documentos multimídia.

Devido à quantidade de dados e informações disponíveis na Internet e disponíveis a qualquer usuário, são exigidas metodologias e ferramentas para extração automática do conhecimento e geração de representações estruturadas a partir desse conhecimento, com o intuito de organizar as informações disponíveis na Web.

Nesse contexto, surgiram diversos métodos, técnicas e ferramentas para o desenvolvimento de ontologias. Com o propósito de auxiliar os engenheiros de ontologias a reduzir tempo e custos no desenvolvimento de novas ontologias.

¹ Folksonomia - é um sistema de classificação derivado da prática e método de criação e gerenciamento colaborativo de *tags* para anotar e categorizar conteúdo. Disponível em: <http://en.wikipedia.org/wiki/Folksonomy>. Acesso em: 25 Mar. 2013.

3.2 Ontologias

Ontologia é um modelo de dados que representa um conjunto de conceitos dentro de um domínio e os relacionamentos entre estes. Ela é utilizada para realizar inferência sobre objetos de um domínio. Um “catálogo de tipo de coisas” em que se supõem existir um domínio.

Uma das definições mais conhecidas para ontologias é apresentada por (GRUBER, 1996):

“Uma ontologia é uma especificação explícita de uma conceitualização. [...] Em tal ontologia, definições associam nomes de entidades no universo do discurso (por exemplo, classes, relações funções etc. com textos que descrevem o que os nomes significam e os axiomas formais que restringem a interpretação e o uso desses termos) [...]” (GRUBER, 1996).

Borst (1997) apresenta uma definição simples e completa: “Uma ontologia é uma especificação formal e explícita de uma conceitualização compartilhada”. Nessa definição, “formal” significa legível para computadores; “especificação explícita” diz respeito a conceitos, propriedades, relações, funções, restrições, axiomas, explicitamente definidos; “compartilhado” quer dizer conhecimento consensual; e “conceitualização” diz respeito a um modelo abstrato de algum fenômeno do mundo real.

3.2.1 Características das ontologias

As ontologias não apresentam sempre a mesma estrutura, mas existem características e componentes básicos comuns presentes em grande parte delas. Mesmo apresentando propriedades distintas, é possível identificar tipos bem definidos. Os componentes básicos de uma ontologia são classes (organizadas em uma taxonomia), relações (representam o tipo de interação entre os conceitos de um domínio), axiomas (usados para modelar sentenças sempre verdadeiras) e instâncias (utilizadas para representar elementos específicos, ou seja, os próprios dados) (NOY; MCGUINNESS, 2001). Para uma melhor compreensão sobre uma ontologia, ver a Figura 4, logo abaixo.

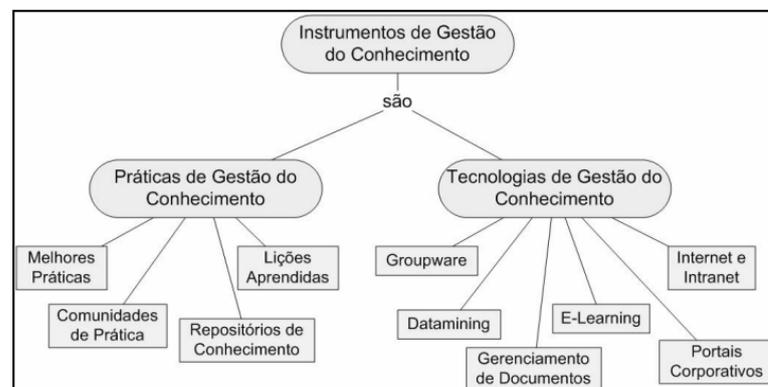


Figura 4 – Exemplo de uma ontologia aplicada a Engenharia do Conhecimento

Fonte: Saito et al. (2007)

3.3 Aprendizado de Ontologias

AO é uma área multidisciplinar que visa à integração de varias áreas de conhecimento para facilitar a construção de ontologias e está relacionada ao processo de aquisição de conhecimento a partir de texto. Ela é composta por um conjunto de métodos e técnicas que são usados na construção de ontologias, enriquecendo-a ou adaptando-a uma ontologia já existente de forma semi-automática. Assim refere-se à extração de elementos (termos, conceitos) de ontologias já existentes ou textos, os quais são utilizados para construir novas ontologias, (BARFOROUSH; SHAMSFARD, 2004). Essa área tem atraído pesquisadores dos mais diversos campos de estudo, tais como: Processamento de Linguagem Natural (PLN), Aprendizagem de Máquinas, Estatística, Linguística, etc.

Outra definição para AO é proposta pelos os autores (SPYNS et al., 2002)(GÓMEZ-PÉREZ; MANZANO-MACHO, 2003) que definem AO como sendo um conjunto de métodos e técnicas usado para construir uma ontologia através de processos automáticos ou semiautomáticos de aquisição de conhecimento via textos, dicionários, bases de conhecimentos, dados semi-estruturados e esquemas relacionados já existentes.

Maedche e Staab (2001) foram os primeiros a descrever sobre o termo AO, como sendo um processo de aquisição de um modelo de domínio².

O objetivo de AO é extrair conceitos e relações relevantes a partir de um dado corpus³ ou outros tipos de conjuntos de dados para formar uma ontologia. Pesquisas sobre o campo de AO possibilitaram o desenvolvimento de várias abordagens que permitem a automação de parte do processo de construção de ontologias (GÓMEZ-PÉREZ; MANZANO-MACHO, 2005) (BUITELAAR et al., 2005).

De acordo com Maedche e Staab (2001) as atuais tecnologias para construção de ontologias (Protégé, Text-to-Onto, OntoEdit) não possibilitam à automatização de todo o processo de construção de ontologias, o que se busca e um processo semi-automático com a menor intervenção humana.

A construção manual de ontologias é uma tarefa que requer muito esforço dos engenheiros de ontologias, seu processo de desenvolvimento requer muito tempo, um alto custo e está propensa a erros. Também considerada uma tarefa exaustiva, não flexível a mudanças e específica para o propósito que foi desenvolvida (BARFOROUSH; SHAMSFARD, 2004)(MAEDCHE, 2002).

Já o processo automático de construção de ontologias resulta em um menor desperdício de tempo, elimina custos e resulta em uma melhor ontologia correspondente com sua aplicação. Tornando-se flexível, pois novas informações podem ser adicionadas ou removidas durante o

² Modelo de domínio é a ontologia que é construída a partir dos textos.

³ Corpus é um conjunto de documentos sobre um determinado tema.

processo de desenvolvimento.

Para se contruir uma ontologia de forma automática ou semi-automática, é necessário uma fonte a partir do qual possa ser extraído conhecimento relevante. Nesse contexto, textos eletrônicos são fontes de informação bastante interessantes, pois a cada dia que passa, mais textos representando os mais diversos domínios estão disponíveis ao acesso. O que torna diferente esses processos de construção de ontologias (manual e automático) é a forma de como são adquiridos os insumos necessários ao seu processo de construção (BASÉGIO, 2008).

3.3.1 Camadas de Aprendizado de Ontologias

A princípio é necessário definir as sub-tarefas (camadas) que constituem a complexa tarefa do aprendizado de ontologias (manual ou com qualquer nível de apoio automatizado). O aprendizado de ontologias está preocupado principalmente com a definição de conceitos e relações entre eles, e com o conhecimento sobre os símbolos que são usados para se referir a eles. Neste caso, torna-se necessário a aquisição de conhecimentos linguísticos sobre os termos usados para se referir a um conceito específico no texto e possíveis sinônimos desses termos. Uma ontologia consiste ainda de uma taxonomia e outras relações não hierárquicas. Finalmente, para que se possam derivar os fatos que não foram explicitamente codificados pela ontologia, é necessária a definição de regras (se possível adquiridas), que permitam tal derivação.

Todos esses aspectos do processo de aprendizado de ontologias podem ser organizados em uma estrutura semelhante a um bolo em camadas, em ordem crescente de complexidade das sub-tarefas, como ilustrado na Figura 5 derivada de (CIMIANO et al., 2005). São apresentadas as camadas e suas diferentes sub-tarefas de aprendizagem de ontologias, descritas abaixo:

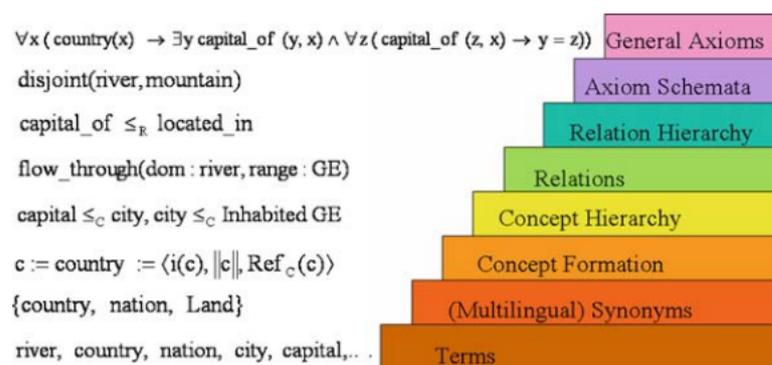


Figura 5 – Camadas de AO

Fonte: (CIMIANO, 2006)

3.3.2 Tarefas do Aprendizado de Ontologias

Nesta seção são descritos todas as tarefas do processo de desenvolvimento de ontologias: *Terms*, *Synonyms*, *Concepts*, *Concept Hierarchy*, *Relations*, *Relation Hierarchy*, *Axiom*

Schemata, General Axioms ilustrados na Figura 5, como demonstrado acima.

Terms

A extração de termos é um pré-requisito para todos os aspectos do processo de aprendizagem de ontologias a partir de textos. Do ponto de vista linguístico, termos são palavras compostas por um significado específico ou técnico sobre um determinado contexto ou domínio. O objetivo dessa tarefa é extrair termos relevantes de conceitos e relações que são característicos para um domínio específico, que será utilizado no processo de AO.

Esta tarefa toma como entrada um conjunto de documentos que representam o domínio de interesse e produz termos (*ontology, ontologies, ontology learning, methodology*) que são utilizados respectivamente para conceitos e relações. A maioria dos métodos utilizados na literatura para extração de termos, que podem ser usados para o aprendizado de ontologias a partir de texto, é baseada em técnicas de recuperação de informação para indexação, e em processamento de linguagem natural (CIMIANO et al., 2005)(CIMIANO, 2006).

Synonyms

A tarefa de identificação de sinônimos consiste em encontrar variações de palavras com significados idênticos ou muito semelhantes, ou seja, encontrar entre os termos de um texto os que compartilham funções semânticas. Duas palavras são consideradas como sinônimos se elas compartilharem um significado em comum, que pode ser utilizado como uma base para formar um conceito relevante para o domínio em questão (FELLBAUM, 1998).

Muito do trabalho desta tarefa está relacionado à integração do WordNet⁴ utilizada para adquirir sinônimos em inglês, e no EuroWordNet⁵ para a aquisição de sinônimos bilíngues ou multilíngues e traduções de termos. Nesta camada é fundamental identificar o sentido apropriado dos termos em questão os quais determinaram o conjunto de sinônimos (*ontology learning, learning ontologies*) que devem ser extraídos.

Segundo Turner (2001) com a crescente importância da web na aquisição de conhecimento a tendência é utilizar medidas estatísticas de informação definidas sobre a web para detectar sinônimos.

Concepts

Ainda existem controvérsias quanto à tarefa de extração de conceitos a partir de texto, pois não está claro o que exatamente constitui um conceito (CIMIANO et al., 2005). Na visão de Buitelaar et al. (2006) uma sugestão ou formação de conceitos deve fornecer:

- Uma definição intencional do conceito (intensão do conceito);

⁴ WordNet é acessível livremente a partir de <http://wordnet.princeton.edu>

⁵ EuroWordNet para mais informações acessar <http://www.illc.uva.nl/EuroWordNet/>

- Um conjunto de instâncias dos conceitos (a sua extensão);
- Um conjunto de realizações linguísticas, ou seja, os termos para esse conceito.

Para finalidade de aprendizagem das ontologias, um conceito pode ser descrito como um par léxico⁶ composto pela intenção do conceito, sua extensão e a descrição das realizações linguísticas, ou seja, os termos (CIMIANO, 2006).

Grande parte das pesquisas em extração de conceitos aborda essa questão a partir de aspectos linguísticos ou textuais. Como alternativa, pesquisadores tem abordado esse problema de um amplo ponto de vista derivando uma hierarquia de entidades a partir dos textos e conceitos visando à descoberta de novos conceitos.

O aprendizado intencional de conceitos inclui a extração ou aquisição formal ou informal de definições. Uma definição formal inclui a extração de propriedades do conceito, sendo essas as relações entre um conceito particular e outros conceitos, tendo como exemplo uma descrição textual.

Concept Hierarchy

Nesta camada são apresentadas as tarefas que constituem a hierarquia de conceitos, propostas por Cimiano (2006) sendo essas: indução, refinamento e ampliação da hierarquia de conceitos.

- Definimos indução hierárquica de conceito como sendo a tarefa de um dado conjunto de conceitos, normalmente junto a suas realizações lexicais formando uma estrutura semi-superior. O seu propósito é induzir uma hierarquia de conceito a partir do zero.
- O refinamento da hierarquia de conceitos é definida com um conjunto de conceitos, bem como uma estrutura semi-superior em pares de aprendizagem. Essa tarefa tem por objetivo estender o conceito de hierarquia já existente com sub-conceitos adicionais de conceitos já existentes, dessa forma ocorre o refinamento da hierarquia.
- A tarefa de extensão lexical ou ampliação da hierarquia de conceitos é definida como um dado conceito em conjunto com a sua referência lexical para se encontrar novas relações lexicais desse conceito, resultando na ampliação lexical.

Na Figura 6 é apresentado um exemplo de hierarquia de conceitos (superclasse – sub-classe) formando uma hierarquia composta por atributos que possuem características comuns entre si.

⁶ Léxico refere-se a um conjunto ou acervo de palavras de um determinado idioma.

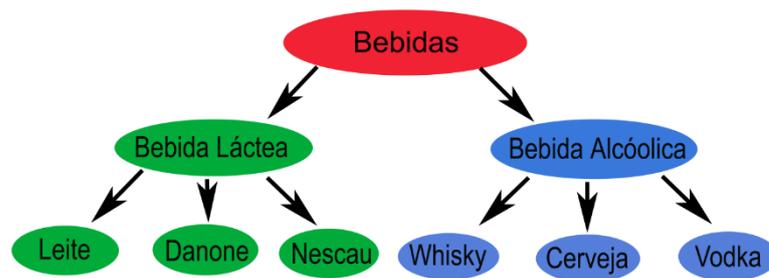


Figura 6 – Hierarquia de Conceitos

Relations (non-hierarchical)

A extração de relações por meio de mineração de texto para o desenvolvimento de ontologias foi introduzido por (MAEDCHE; STAAB, 2000) utilizando regras de associação⁷. A maioria dos trabalhos sobre extração de texto através da mineração de texto⁸ combina a análise estatística com maior ou menor nível de análise linguística explorando a estrutura sintática e dependência para a extração de relações, por exemplo, (BUIBELAAR et al., 2004)(GAMALLO et al., 2002). Trabalhos recentes relacionados à extração de relações a partir de textos têm sido desenvolvidos pelo programa ACE⁹(Extração Automática de Conteúdo).

Axiom Schemata

O objetivo de AO quanto á definição axiomática de conceitos e relações não é aprender o próprio esquema de axiomas, é definir esquemas de axiomas que são frequentemente utilizados em engenharia de ontologias identificando os conceitos, relações ou pares de conceitos. Para saber quais os pares de conceitos que são independentes, quais as relações são simétricas, e a cardinalidade mínima e máxima de uma relação.

General Axioms

Nessa camada o tipo de axioma depende do formalismo lógico que é utilizado. Na tarefa de aprendizagem geral de axiomas, os próprios axiomas têm de ser apreendidos e não apenas instanciados. Os axiomas gerais podem ser pensados como implicações lógicas, restringindo a interpretação dos conceitos e relações. A tarefa de aprender axiomas consiste em derivar relações mais complexas entre conceitos e relações (CIMIANO, 2006). Em axiomas gerais, os próprios axiomas têm de ser apreendidos e não apenas instanciados.

⁷ Regra de associação é uma das tarefas da mineração de dados utilizada para descobrir elementos em comum a um determinado conjunto de dados.

⁸ Mineração de texto refere-se ao processo de aquisição de informações relevantes a partir de texto em linguagem natural.

⁹ ACE está acessível a partir de: <http://www.itl.nist.gov/iad/894.01/tests/ace/>

3.3.3 Métodos para Aprendizado de Ontologias a partir de Texto

Neste tópico são apresentados os métodos utilizados no aprendizado de ontologias, eles utilizam texto como principal fonte de informação para a construção de ontologias ou enriquecimento de ontologias já existentes.

Segundo os autores Gómez-Pérez e Manzano-Macho (2005) os métodos para aprendizagem de ontologias são agrupados em três diferentes categorias: os métodos que aplicam técnicas linguísticas, os baseados em métodos estatísticos e os que se aplicam técnicas de aprendizado de máquina.

Os autores Gómez-Pérez e Manzano-Macho (2005) afirmam que tal classificação é incerta no sentido em que todos os métodos combinam diversas técnicas para atingir aos seus objetivos, portanto, o método utilizado pode pertencer a mais de uma das categorias mencionadas acima. Todos os métodos devem aplicar alguma técnica de processamento linguístico com o objetivo de extrair dos textos partes de informações relevantes para o domínio a qual se destina.

Métodos Baseados em Técnicas Linguísticas

Os métodos ligados a técnicas linguísticas baseiam-se tipicamente em análises sintáticas e morfológicas dos textos para extrair estruturas que servem de base para a construção de ontologias. Um dos métodos mais conhecidos é o de Hearst (HEARST, 1992), ele consiste em padrões linguísticos com a finalidade de detectar as relações taxonômicas (sinônimos, hipônimo¹⁰, etc.) entre palavras ou grupo de palavras.

Os padrões e as medidas de relatividade linguísticas são geralmente construídos com base em características morfológicas ou sintáticas encontradas no texto. Os métodos linguísticos são geralmente aplicados em conjunto com métodos estatísticos com o objetivo de identificar a relevância de cada elemento encontrado nos textos.

Métodos Baseados em Estatística

Este grupo inclui todas as técnicas que baseiam o seu desempenho no cálculo de medidas estatísticas para ajudar os engenheiros de ontologias a detectar novos conceitos ou relações entre eles. As técnicas de análise de frequência de repetição de palavras também chamada de padrões de palavras e a TFIDF¹¹ (*Term Frequency, Inverse Document Frequency*) são aplicadas com o objetivo de identificar os termos (palavras-chaves) contidos em um ou mais textos de acordo

¹⁰ Hipônimo - Uma palavra ou frase cujo alcance semântico (significado) é incluído dentro de outra palavra. Disponível em: <http://pt.wiktionary.org/wiki/hip%C3%B4nimo>. Acesso em: 05 Dez. 2012.

¹¹ TFIDF - (*Term Frequency, Inverse Document Frequency*) é utilizada para indicar a importância de uma palavra ou documento em um conjunto grande e estruturado de texto. Disponível em: <http://en.wikipedia.org/wiki/Tf%E2%80%93idf>. Acesso em: 10 dez 2012.

com a frequência em que elas aparecem, ou seja, o termo que aparecer em maior frequência dentre os documentos será destacado como relevante para a ontologia em domínio. Esse método geralmente é aplicado em conjunto com outras técnicas baseadas em linguística, principalmente processamento de linguagem natural.

Métodos Baseados em Técnicas Aprendizado de Máquina

Neste grupo estão incluídos todos os métodos que tem o seu desempenho fundamentado na utilização de algoritmos de aprendizagem com o intuito de auxiliar os engenheiros de ontologias a detectar novos conceitos ou relações entre eles, e ajudar a encontrar o lugar correto na taxonomia para estes. Esta técnica geralmente é aplicada em conjunto com técnicas de processamento de linguagem natural.

Segundo Fensel et al. (2011) esses métodos podem variar de acordo com algumas entidades ontológicas ou tipos de axiomas a serem adquiridos, como demonstrados a seguir:

- **Conceitos**

São termos (frases nominais ou verbais) referentes a conceitos linguísticos. Essa tarefa tem como objetivo identificar termos ou frases que são relevantes em um domínio particular de interesse. A ocorrência desses termos é determinada comparando-se as frequências estatísticas obtidas para um determinado corpus. Em seguida, com base na hipótese de distribuição definida em Harris (1954) estes termos podem ser agrupados em conjuntos de termos e sinónimos ou através da representação desses termos em um único conceito.

- **Subsunção**

Esse termo foi proposto pelos autores Caraballo (1999) Faure (1999) ele se baseia em técnicas de agrupamento hierárquico (aglomeração e divisão) a fim de organizar de forma hierárquica os grupos de termos linguísticos com comportamentos semelhantes. Segundo (AUSABEL et al., 1978) a subsunção é o processo de ligar novas informações a segmentos de uma estrutura cognitiva preexistente. Ou seja, ele foi concebido para apoiar a aquisição de novos conceitos a partir de outros já existentes, organizando esses conceitos de forma hierárquica formando as hierarquia de conceitos. A vasta maioria destas abordagens dependem dos padrões léxico-sintáticos, conhecidos como padrões de Hearst (HEARST, 1992).

- **Instanciação**

As tarefas de aprendizagem de instanciação, muitas vezes denominado de população da ontologia, são direcionadas para a aquisição de factos no nível de instância das informações. O método para extrair as instanciações e as afirmações de texto em linguagem

natural foi proposto por Cimiano et al. (2005) e é baseado no FrameNet¹² como um tipo específico de conhecimento estruturado.

- **Afirmação**

Seu propósito é identificar as relações entre entidades nomeadas ou indivíduos de uma dada ontologia, ela é geralmente considerada uma subtarefa da extração de informações.

- **Domínio e alcance**

A maioria das abordagens para a aquisição de relações não-taxonômicas baseiam-se na análise de verbos e seus argumentos lexicais. Uma abordagem semi-automática para refinar complexas restrições lógicas de domínio e alcance de propriedades ontológicas foi desenvolvida pelos autores (RUDOLPH, 2008) (RUDOLPH; VOLKER, 2008).

3.3.4 Técnicas para Aprendizado de Ontologias

As técnicas aplicadas a AO estão relacionados diretamente ao tipo dos dados de entrada encontrados em (textos, dicionários, bases de conhecimento e dados semi-estruturados) que são utilizados no processo de construção da ontologia. Navigli e Velardi (2004) Biemann (2005) comentam que os métodos e técnicas aplicadas em AO a partir de texto envolvem áreas de Processamento de Linguagem Natural (PLN) e Aprendizagem de Máquina (AM), sendo um processo extremamente complexo e interdisciplinar. São apresentadas na Tabela 4 algumas das técnicas utilizadas nos respectivos métodos.

Método	Técnicas
PLN	Tokenização, Normalização de frases, Etiquetagem, Padrões gramaticais, Reconhecimento de entidade, Análise de constituintes, co-referência e Detecção automática de sinônimos
AM	k-NN, <i>Naive Bayesian Learning</i> , Árvores de decisão e Métodos de agrupamento hierárquico

Tabela 4 – Técnicas de aprendizado de ontologias a partir de texto

3.3.5 Ferramentas para Aprendizado de Ontologias

As ferramentas de aprendizado de ontologias a partir de texto são utilizadas com a finalidade de apoiar a aquisição ou aperfeiçoamento de axiomas. Essas ferramentas têm como objetivo proporcionar aos engenheiros e desenvolvedores de ontologias o apoio necessário no

¹² FrameNet – para mais informações, acessar: <https://framenet.icsi.berkeley.edu/fndrupal/>

processo de aquisição de conhecimento e conseqüentemente na construção das ontologias de forma automática.

Segundo Fensel (FENSEL et al., 2011), as ferramentas para aprendizado de ontologias são agrupadas em três diferentes tipos:

- No primeiro grupo estão as ferramentas para o aprendizado de novas relações (taxonômicas ou não taxonômicas) a partir de dados de entrada selecionados.
- O segundo abrange todas as ferramentas que auxiliam os engenheiros e desenvolvedores de ontologias a encontrar e estabelecer novos conceitos.
- Finalmente, as ferramentas que auxiliam os engenheiros e desenvolvedores de ontologias a construir uma ontologia ou enriquecer uma já existente.

3.4 Considerações Finais

Aprendizado de Ontologias é um processo utilizado para acelerar a atividade de aquisição de conhecimento do processo de desenvolvimento de ontologias, com o intuito de reduzir esforço e tempo por parte dos engenheiros e desenvolvedores de ontologias na construção de uma ontologia a partir do zero, reutilização de uma já existente ou acelerar a construção de ontologias a ser utilizada para diferentes fins. No entanto, as atuais tecnologias não possibilitam a automatização de todo o processo de construção de ontologias. Em vista dos fatos, vários métodos e ferramentas têm sido propostos para acelerar este processo usando diferentes fontes de informação e diferentes técnicas.

4 Metodologia

A metodologia científica é fundamental para guiar o desenvolvimento de um trabalho acadêmico, nela é apresentada a estrutura a ser seguida para se obter uma melhor produtividade e qualidade dos trabalhos. Ela traz credibilidade ao estudo e torna possível sua replicação, de forma independente, por outros pesquisadores. Este capítulo apresenta a abordagem metodológica usada nesta pesquisa e está organizado em duas partes, como descritos na Tabela 5 a seguir:

Abordagem	É apresentado
Classificação da Pesquisa	- Classificação da Pesquisa - Quadro metodológico
Ciclo da Pesquisa	- Etapas da Pesquisa - Processo do Mapeamento Sistemático

Tabela 5 – Tipos de abordagem metodológica

4.1 Classificação da Pesquisa

Esta pesquisa utiliza o método de abordagem Experimental baseado em dados de natureza qualitativa. O método de procedimento utilizado será o Estudo de Mapeamento Sistemático (EMS) que é um tipo de revisão sistemática. Na Tabela 6, a seguir é apresentado o quadro metodológico da pesquisa.

Quadro Metodológico	
Método de Abordagem	- Indutivo (Científico)
Método de Procedimento	- Estudo de Mapeamento Sistemático
Natureza dos Dados	- Qualitativa

Tabela 6 – Quadro Metodológico da Pesquisa

É um método responsável pela generalização, onde parte-se de algo particular para uma questão mais ampla, ou seja, geral. Para Marconi e Lakatos (2007), indução é um processo mental por intermédio do qual, partindo de dados particulares, suficientemente constatados, infere-se uma verdade geral ou universal, não contida nas partes examinadas.

Quanto ao método de procedimento utilizado, que é a etapa mais concretano processo

de busca, foi definido para esta pesquisa o Estudo de mapeamento Sistemático (EMS), do inglês *Systematic Mapping Study*, que é uma forma de avaliar e interpretar todas as pesquisas disponíveis referentes a uma questão de pesquisa em particular, área temática, ou fenômeno de interesse.

Por fim, no que diz respeito à utilização da pesquisa qualitativa, os autores Marconi e Lakatos (2007) destacam algumas vantagens e desvantagens quanto a sua utilização se comparando aos demais métodos (quantitativo, estatístico). A pesquisa qualitativa se preocupa em interpretar e analisar os aspectos mais profundos, fornecendo análises mais detalhadas, além de prover informações mais exploratórias e auxiliar a refinar as preposições para um melhor ajuste dos dados. No entanto, a utilização de pesquisa qualitativa traz algumas desvantagens. Dentre elas, exige um maior esforço que o pesquisador tem que despender em relação à pesquisa quantitativa e uma maior dificuldade para resumir os achados, e seus resultados são considerados mais nebulosos do que os quantitativos.

4.1.1 Classificação segundo Cooper

Com o intuito de complementar a metodologia apresentada, será utilizada a taxonomia sugerida por (COOPER, 1988) utilizada para classificar revisões sistemáticas. Para tal classificação, o autor propõe seis características: foco, objetivo, perspectiva, cobertura, organização e audiência. Esta classificação é apresentada na Tabela 7:

Características	Descrição
Foco	- Resultado de Pesquisas - Métodos de Pesquisa - Práticas ou Aplicações
Objetivo	- Representação Neutra - Exposição de Posicionamento
Perspectiva	- Integração - Análise Crítica da Literatura
Cobertura	- Exaustiva - Exaustiva com Seleção de Citação - Representativa e Central - Essencial
Organização	- Conceitual - Metodologicamente
Audiência	- Pesquisadores Especialistas - Pesquisadores em Geral - Profissionais

Tabela 7 – Classificação segundo Cooper

Cada etapa da classificação de Cooper (1988) serão descritas a seguir:

- **Foco:** está etapa é relacionado ao interesse principal do pesquisador, sendo composta pelas seguintes categorias: Resultados da Pesquisa, Métodos de Pesquisa e Práticas ou aplicações. Uma revisão geralmente é composta por mais de um foco, cada qual variando o seu nível de atenção.
- **Objetivo:** foca nos resultados que o pesquisador deseja alcançar. Estes resultados são classificados como: Integração, Crítica e Identificação de Problemas Centrais. O objetivo mais comum encontrado em uma revisão é o da integração, pois corresponde à síntese da literatura existente em uma área. Também pode ser uma análise crítica da literatura, nesse caso a intenção do pesquisador é demonstrar que os resultados dos estudos anteriores são injustificáveis e por fim tem a identificação de problemas centrais, com a intenção de fornecer sugestões sobre como problemas de uma área se tornaram evidentes.
- **Perspectiva:** são as discussões das posições assumidas pesquisas pelos pesquisadores. Subdividindo-se em duas nas categorias: Representação Neutra ou exposição de conhecimento. Na primeira perspectiva, a intenção do pesquisador é evidenciar sobre os diferentes métodos, técnicas e ferramentas de forma a avaliar os trabalhos produzidos na literatura. Na segunda perspectiva, o pesquisador tem a intenção de apresentar e defender o seu ponto de vista.
- **Cobertura:** é exposta a decisão tomada pelo pesquisador sobre como realizar as bus-

cas e a inclusão dos trabalhos relevantes para sua pesquisa, ou seja, como o revisor vai pesquisar a literatura e a sua decisão sobre a qualidade do material. São quatro tipos de cobertura: Exaustiva, Exaustiva com seleção de citação, Representativa e Central ou Essencial. Uma breve descrição sobre os tipos de cobertura é apresentada na Tabela 8, a seguir:

Tipo de cobertura	A intenção do pesquisador
Exaustiva	- Coletar toda ou quase toda a literatura disponível na área
Exaustiva com seleção de citação	- Apresentar apenas uma amostra de todo o trabalho selecionado
Representativa	- Apresentar uma amostra que representa toda a área de pesquisa
Central ou Essencial	- Selecionar os trabalhos mais relevantes que direcionam o campo da pesquisa

Tabela 8 – Descrição da cobertura do MS

- *Organização*: diz respeito à organização dos trabalhos e dos resultados. Podendo ser apresentados em três diferentes formas: Historicamente, Conceitualmente, Metodologicamente. Pesquisadores que realizam a revisão podem combinar mais de uma forma de organização do seu trabalho. Uma breve descrição sobre esses tipos de organização e apresentado na Tabela 9, a seguir.

Forma de Organização	Descrição
Historicamente	- Os temas são apresentados em ordem cronológica do aparecimento na literatura
Conceitualmente	- São apresentados os trabalhos relacionados que apresentam as mesmas ideias
Metodologicamente	- Os trabalhos com métodos semelhantes são agrupados em subtemas

Tabela 9 – Organização dos trabalhos e resultados

- *Audiência*: nessa característica é apresentado o público alvo a que se destina a pesquisa (Pesquisadores em Geral, Engenheiros de Ontologias, Desenvolvedores de Ontologias e usuários interessados em seu desenvolvimento).

4.2 Ciclo da Pesquisa

As etapas que compõem o ciclo de vida utilizado nesta pesquisa são apresentadas na Figura 7. O problema a ser pesquisado está relacionado à área de Aprendizado de Ontologias,

onde se pretende evidenciar através da execução de um mapeamento sistemático da área, os métodos, técnicas e ferramentas usadas no seu aprendizado. Em seguida, o foco desta pesquisa será definido em uma questão de pesquisa que será apresentada mais adiante.

Posteriormente, será desenvolvido e executado um protocolo para a execução do Estudo de Mapeamento Sistemático (EMS) com intuito de coletar evidências para responder a questão de pesquisa proposta. Em seguida, os estudos coletados serão selecionados de acordo com a questão de pesquisa a ser respondida. E o passo seguinte é a extração das informações.

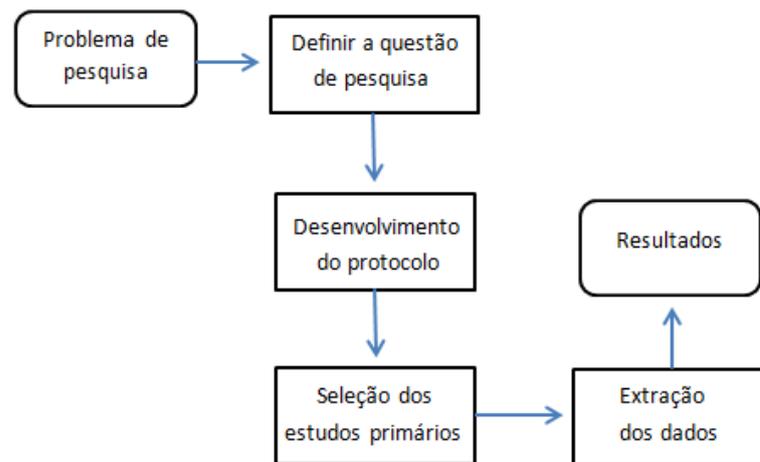


Figura 7 – Etapas de Pesquisa

4.2.1 Mapeamento Sistemático

Neste tópico são apresentados os pontos seguidos pelo mapeamento sistemático com a intenção de encontrar e analisar o maior número de trabalhos relevantes à área de aprendizado de ontologias a fim de responder a questão de pesquisa, definida mais adiante na seção 4.2.2.

Segundo (CHARTERS; KITCHENHAM, 2007) as perguntas de pesquisa de um EMS são mais amplas e gerais (normalmente de carácter exploratório), em contraste com que são elaboradas em Revisões Sistemáticas da Literatura, que devem ser bem mais precisas.

Segundo (KITCHENHAM et al., 2004) revisões sistemáticas e mapeamento sistemático de estudos possuem em comum:

- **Identificação da necessidade de se executar revisão sistemática;**

Antes de se realizar uma RSL ou MSE é importante saber qual a necessidade da sua realização e definir uma área na qual vai ser realizado.

- **Elaboração da questão de pesquisa;**

Nessa etapa é definida a questão de pesquisa na qual se deseja responder com os resultados obtidos a partir da execução da RSL ou MSL;

- **Uma abrangente e exaustiva busca por estudos primários;**
È realizada a busca por estudos relevantes que possam responder a questão de pesquisa.
- **Avaliação de qualidade dos estudos incluídos;**
Nessa fase é realizada uma análise para averiguar a qualidade dos documentos adquiridos, que foram incluídos no estudo.
- **Identificação dos dados necessários para responder as questões da pesquisa;**
Nessa etapa os documentos são agrupados de acordo com cada questão de pesquisa.
- **Extração dos dados;**
Nessa acontece o processo de extração de dados dos documentos que foram selecionados, obtendo-se assim as informações necessárias sobre as características do estudo e os resultados dos estudos incluídos.
- **Resumo e síntese dos resultados do estudo (meta-análise);**
Nessa etapa as informações são resumidas em dados e são geralmente armazenados em planilhas eletrônicas para facilitar a sua manipulação.

4.2.2 Escopo e Questão de Pesquisa

Com o objetivo de traçar o escopo desta pesquisa para identificar os elementos que vieram a fazer parte da questão de pesquisa, foi utilizada a estrutura citada por (CHARTERS; KITCHENHAM, 2007). Os autores (PETTCRIEW; ROBERTS, 2008) recomendam desenvolver as questões de pesquisa a partir da estrutura ao qual ele denomina de **PICOC**:

- **População (*Population*):** Engenheiros de Ontologias, desenvolvedores de Ontologias, usuários interessados em seu desenvolvimento.
- **Intervenção (*Intervention*):** Mecanismos automáticos de construção de ontologias.
- **Comparação (*Comparasion*):** Não é aplicada.
- **Resultado (*Outcomes*):** Mapeamento sistemático da Literatura (MSL).
- **Contexto (*Context*):** Indústria é academia.

O item da estrutura denominado Comparação (*Comparasion*) não será utilizado, uma vez em que o estudo não realiza comparações entre os métodos, técnicas e ferramentas utilizados no aprendizado de ontologias. O item Contexto (*Context*) é utilizado para definir o contexto ao qual ocorre a comparação. Entretanto, aqui ele será utilizado para definir o contexto ao

qual se aplicam os resultados encontrados. Após definir a estrutura **PICOC** pergunta que guiou estudo foi definida.

- **QP1: Quais são os métodos utilizados no Aprendizado de Ontologias?**

Nesse ponto, o mapeamento sistemático (MS) pretende responder quais os métodos existentes para o aprendizado de ontologias.

- **QP2: Quais são as técnicas utilizados no Aprendizado de Ontologias?**

Nesse ponto, o mapeamento sistemático (MS) pretende responder quais as técnicas existentes para o aprendizado de ontologias.

- **QP3: Quais são as ferramentas utilizados no Aprendizado de Ontologias?**

Nesse ponto, o mapeamento sistemático (MS) pretende responder quais os ferramentas existentes para o aprendizado de ontologias.

4.2.3 Estratégia de Busca

String de Busca

A principal parte de estratégia de busca é a elaboração da *string* de busca, que é construída com base em termos selecionados da questão de pesquisa. A construção da *string* de busca seguiu uma estratégia baseada em (KITCHENHAM et al., 2006), que consiste nos seguintes passos:

- Selecionar a partir da questão de pesquisa os termos para construir a *string*;
- Realizar a tradução desses termos para o inglês, por ser o idioma mais utilizado na literatura da Ciência da Computação.
- Identificar os sinônimos referentes aos termos selecionados na questão de pesquisa;
- A *string* de busca é gerada a partir da combinação desses termos principais e seus sinônimos;
- Utilizar o conectivo booleano *OR* para interligar os sinônimos;
- Utilizar o conectivo booleano *AND* para interligar os termos.
- Verificar a *string* de busca construída realizando testes nas bibliotecas selecionadas e comparando a relevância dos documentos obtidos;

Termos	Sinônimos
<i>ontology learning</i>	
<i>Method</i>	<i>methodology, process</i>
<i>Technique</i>	<i>techniques</i>
<i>Tool</i>	<i>framework, plug-in</i>

Tabela 10 – Termos e Sinônimos

Com base na estratégia descrita na Tabela 10, foi obtida a seguinte *string* de busca:

("ontology learning") AND (framework OR method OR methodology OR process OR technique OR thecniques) AND (tool OR plugin)

Listagem de Fontes

Baseado no tópico Fontes de Busca descrito acima e com o auxílio de especialista sobre o assunto, foram selecionadas as seguintes fontes de buscas automáticas e manuais para a realização do presente trabalho, como demonstrados a seguir:

Base Eletrônica	URL
<i>IEEEExplore Digital Library</i>	http://ieeexplore.ieee.org/
<i>ACM Digital Library</i>	http://portal.acm.org
<i>SciencDirect</i>	http://www.sciencedirect.com
<i>Elsevier Scopus</i>	http://www.scopus.com

Tabela 11 – Fontes para busca automática

Fonte	Base Indexadora
<i>Semantic Web Journal</i>	
<i>Journal of Web Semantic</i>	<i>Elsevier SciencDirect</i>

Tabela 12 – Fontes para busca manual (Journals)

4.2.4 Critérios de Inclusão e Exclusão e Procedimentos de Seleção dos estudos no MSL

Critérios de Inclusão

- **(Restrição 1)** Estudos que estão de acordo com a questão de pesquisa a ser analisada.
- **(Restrição 2)** Se dois artigos publicarem resultados diferentes de um mesmo estudo, ambos serão incluídos.

Critérios de Inclusão

Serão excluídos:

- **(Restrição 3)** Artigos que não estejam escritos em inglês;
- **(Restrição 4)** Estudos completamente irrelevantes para a pesquisa, ou seja, os que não responde a questão de pesquisa definida;
- **(Restrição 5)** Artigos que não estejam disponíveis para recuperação através da web;
- **(Restrição 6)** Estudos repetidos, se um estudo estiver disponível em diferentes fontes, a primeira pesquisa será considerada.
- **(Restrição 7)** Se dois artigos diferentes publicarem os mesmos resultados de um estudo, o menos detalhado será excluído;
- **(Restrição 8)** Estudos duplicados, caso exista dois um mais estudos iguais de uma mesma fonte, um deles será excluído.
- **(Restrição 9)** Estudos que apresentam textos, conteúdo e resultados incompletos.
- **(Restrição 10)** Artigos que não sejam da área de Ciência da Computação (por exemplo, Administração, Medicina, etc.);
- **(Restrição 11)** Os artigos cujo ano de publicação seja inferior a 2001(ano em que ocorreu a publicação marcante sobre o tema *Ontology Learning*: a obra *Ontology Learning for the Semantic Web*, dos pesquisadores Alexander Maedche e Steffen Staab (MAEDCHE; STAAB, 2001).

Processo de seleção dos dados

Nessa etapa serão realizadas as buscas de acordo com a estratégia de busca descrita nas seções anteriores. Onde a partir da leitura do título, *abstract* e palavras-chaves dos estudos, excluem trabalhos utilizando os critérios de exclusão de **Restrição 3, 4, 5, 6, 7, 8, 9, 10, 11**. Caso tenha duvida de que o artigo não atenda a estes critérios e exclusão, o mesmo deve ser incluído para ser avaliado na próxima fase.

O passo seguinte é formar uma lista de potenciais estudos primários, utilizando-se ainda o critério de exclusão de Restrição 8. Essa lista é mantida no sistema de gerenciador de compartilhamento **Dropbox** (<https://www.dropbox.com/>) Também será utilizado o **Mendeley** (<http://www.mendeley.com/>) como um banco de dados dos estudos. Esta ferramenta também

extrai os dados de cada estudo primário. Os dados extraídos da ferramenta **Mendeley** serão exportados para uma planilha do *Microsoft Excel* contendo as seguintes colunas, como descritos na Tabela 13, a seguir:

Informação	Descrição
<i>Id Estudo</i>	Identificador do estudo
Título	O título do estudo
Autor (es)	Lista de autores do estudo
Ano	Ano de publicação do estudo
País	Lista de países das instituições de origem dos autores
Fonte	Fonte onde o estudo foi encontrado

Tabela 13 – Informações coletadas dos estudos primários

O próximo passo é realizar a seleção, onde cada estudo da lista de potenciais estudos será avaliado, mediante a leitura da Introdução e Conclusão. Caso haja dúvida se o artigo responde ou não a questão de pesquisa, outras seções do artigo devem ser lidas. Utilizam-se os critérios de exclusão Restrição 4, 7, 10 e os critérios de inclusão Restrição 1, 2 para se chegar a uma lista final dos estudos primários, os quais muitos provavelmente respondem a questão de pesquisa. Toda a seleção será documentada em uma planilha do *Microsoft Excel*, contendo as colunas especificadas na Tabela 14 demonstrados a seguir:

Informação	Descrição
<i>Id Estudo</i>	Identificador do estudo
Título	O título do estudo
Autor (es)	Lista de autores do estudo
Ano	Ano de publicação do estudo
País	Lista de países das instituições de origem dos autores
Fonte	Fonte onde o estudo foi encontrado
Pergunta de Pesquisa	A pergunta de pesquisa que se pretende responder
Critério de Exclusão	Restrição pelo qual o estudo foi excluído do mapeamento

Tabela 14 – Informações coletadas dos estudos selecionados

4.3 Considerações Finais

No presente capítulo foi apresentado à metodologia utilizada para realização desta pesquisa, demonstrando a sua estrutura e o processo de condução. Descrevendo o o protocolo utilizado para nortear a execução do EMS proposto pelo estudo.

5 Resultados

Este Capítulo apresenta os resultados do Estudo de Mapeamento Sistemático obtidos através da execução do protocolo definido no capítulo anterior. Os resultados estão estruturados em três componentes, descritos a seguir.

- Extração e análise dos dados: apresentação dos dados gerais sobre o estudo, tais como: quantidade de estudos retornados na busca automatizada, processo de seleção com o número final de estudos primários, a distribuição das datas de publicação de cada um ao longo dos anos, dentre outros;
- Mapeamento das evidências: nesta etapa são apresentados trechos copiados dos estudos primários selecionados, que foram coletados como evidências, ou seja, os resultados para as questões de pesquisas QP1, QP2, QP3;
- Discussão sobre os resultados: apresentação dos principais resultados obtidos com o estudo.

5.1 Extração e Análise dos Dados

O mapeamento sistemático de estudos foi realizado de acordo com o protocolo apresentado no Capítulo 4. A *string* de busca definida juntamente com as quatro bibliotecas digitais usadas no estudo, retornaram 1931 trabalhos, sendo, 668 resultados da busca na *IEEE*, 346 na *ACM*, 239 na *ScienceDirect* e 678 na *Scopus*. O gráfico na Figura 8 ilustra a participação das bibliotecas digitais no montante de resultados encontrados.

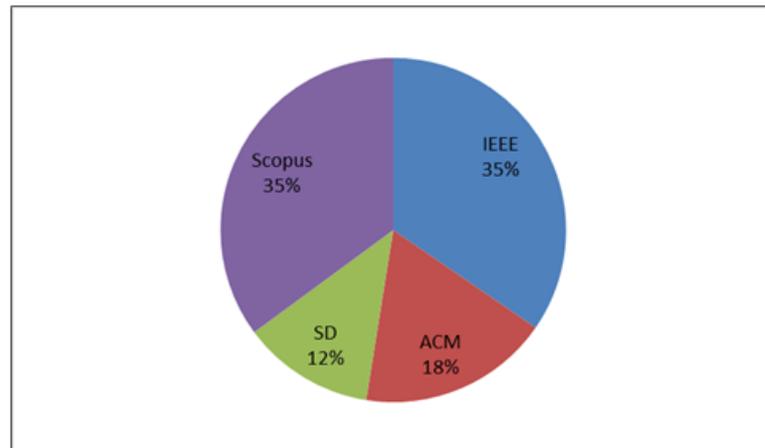


Figura 8 – Resultados da busca automatizada.

Além da busca realizada nas bibliotecas digitais, buscas manuais também foram empreendidas com objetivo de aumentar o alcance do estudo. As buscas manuais foram realizadas em *Journals* de relevância para a área estudada, foram eles: *Semantic Web Journal (SWJ)* e *Journal of Web Semantic (JSW)* no período de 2001 a 2013. O montante de estudos que cada *Journal* tem a oferecer é conhecido através de uma simples contagem de estudos relevantes para a área estudada, o resultado da contagem foi 47 trabalhos, dos quais 24 estudos são do *SWJ* e 23 do *JSW*. O gráfico na Figura 9 ilustra a participação de cada *Journal* no montante de estudos encontrados na busca manual.

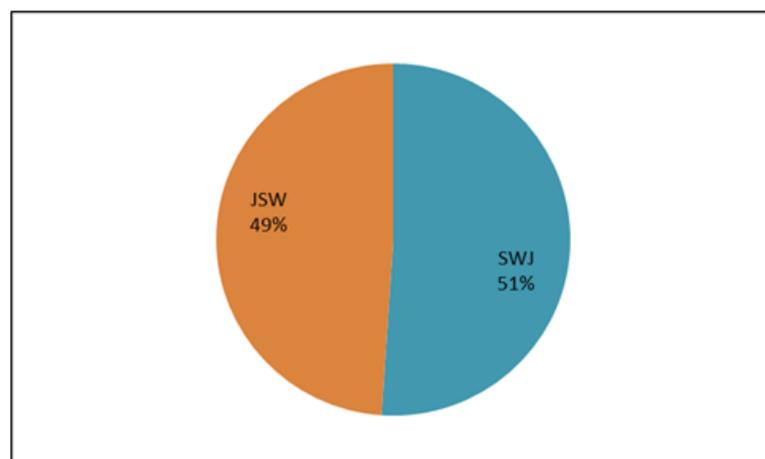


Figura 9 – Resultados da busca manual.

O resultado final dos dois tipos de busca apresentou um espaço de busca de estudos primários cujo montante é de 1978 estudos, onde aproximadamente 2% é o resultado da busca manual e 98% da busca automatizada. O gráfico na Figura 10 ilustra esse resultado.

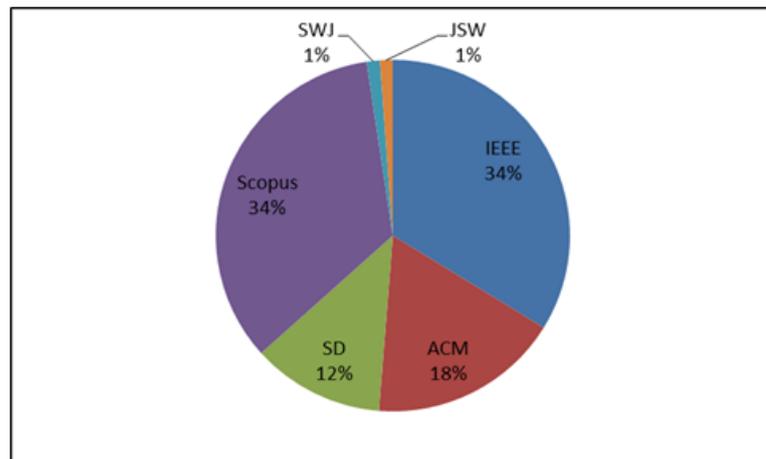


Figura 10 – Resultados da busca automatizada e manual.

Como pode ser observado, o número de estudos retornados pela busca automatizada e manual foi alto. Este problema é compartilhado pelos estudos secundários sistemáticos (revisões sistemáticas da literatura e mapeamentos sistemáticos de estudos) que utilizam processos automatizados de busca em função das características e funcionalidades dos engines de buscas (CHARTERS; KITCHENHAM, 2007).

Após a leitura do título, resumo, introdução, conclusão e aplicação dos critérios de inclusão e exclusão apresentados no Capítulo 4, o alto número de estudos retornados inicialmente, foi reduzido de 1978 para 43 estudos primários disponíveis no apêndice B, (uma redução de aproximadamente 97,8%). Dessa forma os 1935 estudos foram excluídos de acordo com os critérios de Restrição pré-definidos no Capítulo 4, por serem repetidos/duplicados, eram irrelevantes, não eram da área de Ciência Computação, não eram artigos completos, não estarem escritos em inglês ou simplesmente não se teve acesso ao artigo. O Gráfico da Figura 11 mostra a distribuição dos estudos primários selecionados.

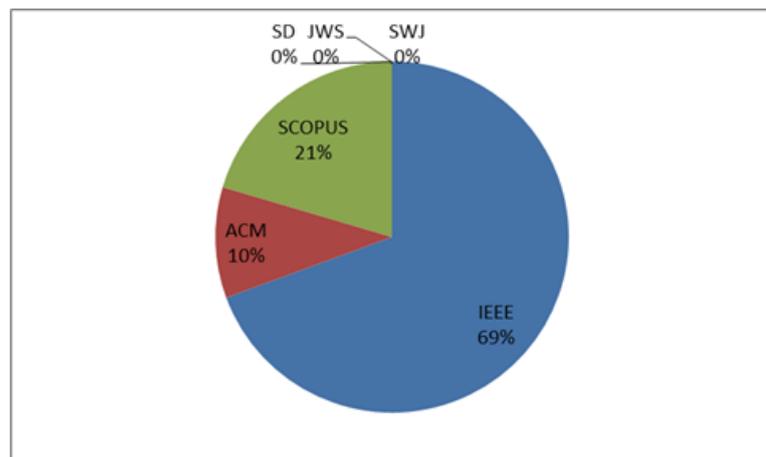


Figura 11 – Resultado seleção de trabalhos primários que compõem o estudo.

Uma observação interessante pode ser feita em relação à quantidade relevante de estudos repetidos pela *SciencDirect* chegando a um montante de 18% dos estudos. Isso pode ser explicado pelo fato de a *SciencDirect* ter sido o último engenho de busca a ser utilizado e consequentemente seus estudos foram os últimos a serem analisados. Outro fato importante a ser observar está comumente relacionado à quantidade de estudos não disponíveis para acesso pelos engenhos de busca da *Scopus* com 55% e *ACM* 33% dos estudos. A Figura 12 mostra de forma sintética a evolução do processo de seleção dos estudos primários.

Seleção dos estudos primários								
FONTES	Excluídos						Incluídos	
	Restrição 3	Restrição 4	Restrição 5	Restrição 6	Restrição 8	Restrição 9	Restrição 6	Estudos primários
<i>IEEE</i>		628	5		6			29
<i>ACM</i>	1	166	168	1	1	3	1	5
<i>SCOPUS</i>		120	547		2			9
<i>SCIENC DIRECT</i>		173	11	54	1			0
<i>JSW</i>		18		5				0
<i>SWJ</i>		24						0

Figura 12 – Evolução do processo de seleção dos estudos.

As buscas realizadas (automatizada e manual) foram limitadas a um período específico, todos os estudos encontrados estão entre 2001 e 2013, sendo a sua maioria concentrado na última década. Isso evidencia que pesquisas na área de Aprendizado de Ontologias têm aumentado na última década. O gráfico da Figura 13 ilustra a concentração dos estudos por ano.

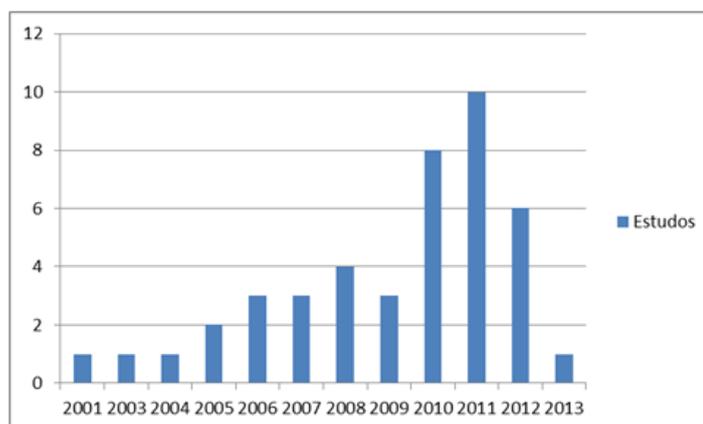


Figura 13 – Distribuição dos estudos ao longo dos anos.

O mapeamento contabilizou 142 autores nos 43 estudos primários selecionados, dentre os quais estão presentes em mais de um artigo: Nathalie Aussenac-Gilles, Paola Velardi, Sahar Changuel, Ricardo Gacitua, Alexander Maedche, Bill Richardson, Roberto Navigli. Os pesquisadores tem origem de 20 países diferentes conforme detalhamento na Figura 14.

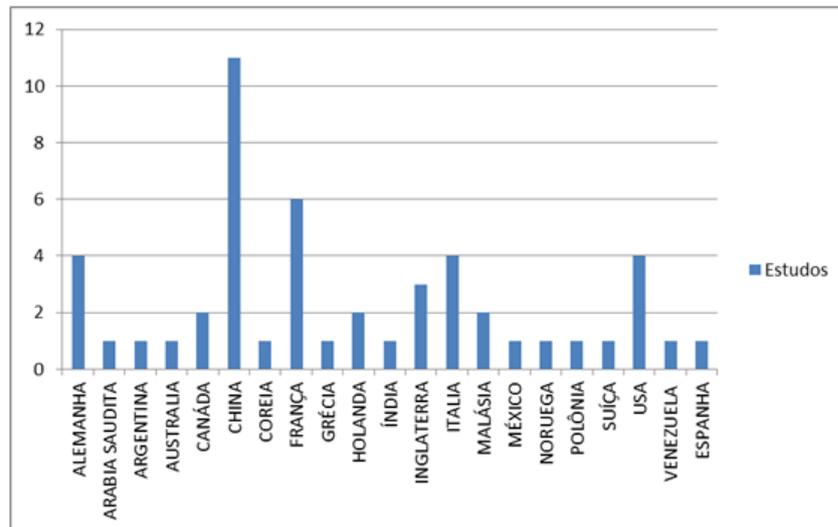


Figura 14 – Distribuição dos estudos pelos países das instituições dos pesquisadores.

A soma das publicações de cada país supera a quantidade de estudos selecionados porque alguns estudos foram desenvolvidos em cooperação entre dois ou mais pesquisadores que tem origens em instituições de países diferentes.

Esta seção apresentou dados gerais coletados no mapeamento sistemático. Tais informações podem ser usadas como referência para estudos semelhantes ou replicações para confirmar ou refutar os dados apresentados.

5.2 Mapeamento das Evidências

Nesta seção, as evidências capturadas através dos estudos primários selecionados são apresentadas respondendo as perguntas de questão de pesquisa que guiaram o estudo. A seção é dividida em três partes, cada uma respondendo a uma questão de pesquisa. A Seção 5.2.1 apresenta as evidências quanto aos métodos utilizados no aprendizado de ontologias. Na Seção 5.2.2 são apresentadas as evidências quanto às técnicas usadas em aprendizado de ontologias, só ressaltando que essas técnicas são empregadas nos métodos e conseqüentemente são utilizados para desenvolvimento das ferramentas. E por fim, na Seção 5.2.3 são apresentadas as ferramentas utilizadas o aprendizado de ontologias. Todas as evidências são apresentadas com a devida referência. Os Estudos Primários são referenciados com o prefixo EP mais a numeração de 1 a 43 na ordem em que são dispostos na lista no Apêndice B.

5.2.1 Métodos

QP1 – Quais são os métodos utilizados no aprendizado de ontologias?

O principal objetivo dessa questão é evidenciar quais os métodos utilizados no aprendizado de ontologias. Dos 43 estudos primários pertencentes ao mapeamento, foram detectadas 5 diferentes métodos que são aplicados no aprendizado de ontologias. Uma breve descrição desses métodos é apresentada na Tabela 15, a seguir:

Método	Descrição	EP
<i>TSD</i>	<i>Term Sense Disambiguation</i> é um método aplicado a aprendizado de ontologias com intuito de identificar informações contextuais relevantes da <i>Web</i> .	EP1
OntOAIr	OntOAIr um método de aprendizagem de ontologias para a construção de ontologias leves a partir de várias coleções de documentos.	EP12
RTAXON	RTAXON é um método de aprendizagem de ontologias a partir de bancos de dados relacionais, e abordagens de mineração de dados.	EP23
OntoLearn	OntoLearn é um método de aprendizagem ontologia que aplica um algoritmo hierárquico a um conjunto de documentos. Este método propõe uma interpretação semântica de termos a fim de identificar conceitos compostos por duas ou mais palavras.	EP12
SMOL	Systemic Methodology for OL é uma método baseado em sistemas que combinam recursos metodológicos (MR) para OL.	EP2

Tabela 15 – Métodos para aprendizado de ontologias

5.2.2 Técnicas

QP2 – Quais são as técnicas utilizadas no aprendizado de ontologias?

O objetivo após dessa questão é mapear as técnicas utilizadas no aprendizado de ontologias. Nos 43 estudos analisados, foram encontradas 24 técnicas. Essas técnicas estão classificadas de acordo com as seguintes abordagens: baseadas em Estatística, baseadas em Linguística e baseadas em Lógica. Algumas técnicas foram referenciadas com maior frequência dentre o montante de técnicas que foram detectadas nos estudos primários. Na Tabela 16 é mostrada uma breve descrição sobre essas técnicas.

Técnica	Tipo	Descrição	EP
RAI	Linguística	<i>Relevance-driven Abstraction Identification</i> é uma técnica para identificação de abstrações simples e multi-palavras.	EP5
<i>Agglomerative Clustering</i>	Estatística	Agglomerative Clustering é uma técnica usada para agrupar conceitos individuais, agrupando ou relacionando esses conceitos.	EP30
<i>Co-occurrence analysis</i>	Estatística	Técnica utilizada para identificar a co-ocorrência de termos presentes em documentos, a fim de extrair termos relacionados.	EP30
<i>Association Rule mining</i>	Estatística	Técnica utilizada para descrever as associações entre os conceitos de um dado conjunto de pares de documentos.	EP30, EP37
<i>Sentence parsing</i>	Linguística	A técnica de análise de sentença é utilizada para analisar um conjunto de documentos a fim de descobrir os termos e suas relações presentes.	EP30
<i>Syntactic structure analysis</i>	Linguística	Essa técnica examina as informações sintáticas para descobrir termos e suas relações em frases.	EP2
<i>Subcategorization frames</i>	Linguística	Técnica utilizada para analisar palavras e estruturas sintáticas (substantivos, expressões verbais e frases preposicionais) para descobrir termos e suas potenciais relações.	EP30
<i>Part-of-speech tagging</i>	Linguística	Técnica utilizada para marcar uma palavra em um texto, a fim de identificar sua relação com outras palavras adjacentes e relacionadas em uma frase ou parágrafo.	EP30
<i>Lexico-syntactic patterns</i>	Linguística	Técnica utilizada para extrair relações entre palavras.	EP30
<i>Inductive logic programming</i>	Lógica	Técnica utilizada para derivar regras a partir de uma coleção existente de conceitos e relações positivas e negativas.	EP30
<i>logical inference</i>	Lógica	Técnica utilizada para derivar relações implícitas por meio de regras, tais como herança e transitividade. Usando o exemplo clássico, dado que "Sócrates é um homem" e "Todos os homens são mortais", podemos descobrir uma nova relação atributo afirmando que "Sócrates é mortal".	EP30
<i>term subsumption</i>	Estatística	Técnica utilizada para detectar as probabilidades de ocorrência de termos em documentos a fim de descobrir relações hierárquicas entre eles. Ela é usada para quantificar a extensão de um termo X ser mais gerado que um termo Y.	EP37

Tabela 16 – Técnicas para aprendizado de ontologias

Além das técnicas mencionadas acima, foram detectadas outras técnicas dentre os estudos primários analisados: as baseadas em Estatísticas (*semantic analysis, contrastive analysis, Formal Concept analysis, information retrieval*), baseadas em Linguísticas (*Anaphora resolution, Use of semantic templates and domain knowledge, Concepts and glossary from WordNet, Use of domain léxicon Word sence desambiguation, Shallow parsing, Relevance analysis, Part-of-speech tagging using WordNet, Hipernyms from WordNet*).

5.2.3 Ferramentas

QP3 – Quais são as ferramentas utilizadas no aprendizado de ontologias?

O objetivo dessa questão é mapear as ferramentas utilizadas no aprendizado de ontologias. Nos 43 estudos analisados, foram encontradas 19 ferramentas que são utilizados no aprendizado de ontologias. Algumas ferramentas foram referenciadas com maior frequência nos estudos primários, em relação ao conjunto de ferramentas que foram detectadas com a execução do mapeamento. Na Tabela 17 é mostrada uma breve descrição sobre essas ferramentas.

Além das ferramentas mencionadas na Tabela 17, outras ferramentas foram referenciadas com uma menor frequência dentre os estudos que compõem o mapeamento sistemático, sendo: *OntoBuilder, OntoGain, OntoEdit, OntoLiFT, OntoMiner, OntoCmaps, Semi-Automatic Domain Ontology Acquisition Tool (SOAT), Developing Ontology-Grounded Methods and Applications (DOGMA), Information System for Ontology Learning and Domain (ISOLDE), ONtologies In ONtology (ONION), Concept-Relation-Concept Tuple-based Ontology Learning (CTRCTOL)*.

Ferramentas	Descrição	EP
ASIUM	É um sistema semi-automático para aprendizado de ontologias baseado em aquisição de conhecimento. O objetivo é ajudar o especialista na aquisição de conhecimento semântico a partir de textos por meio de análise sintática.	EP3, EP13, EP30, EP42
<i>TextStorm/Clouds</i>	É um sistema semi-automatizado de aprendizado de ontologias utilizado para construção semi-automática de uma rede semântica que contém apenas os conceitos e suas relações, através de um texto relevante para o domínio alvo.	EP3, EP30
SYNDIKATE	É um sistema de aquisição automática de conhecimento a partir de textos usado para transferir seus conteúdos para estruturas de representação formal que constituem uma base de conhecimento. Caracterizado pela aprendizagem incremental de conceitos e relações baseadas em compreensão de texto, usando linguística.	EP3, EP30
<i>OntoLearn</i>	Ferramenta para aprendizado de ontologias utilizada para extrair termos relevantes a partir de corpos de domínio de texto (<i>web sites</i>), relaciona-los com os conceitos apropriados de uma ontologia de propósito geral, e para detectar as relações entre esses conceitos.	EP3, EP4, EP11, EP13, EP16, EP17, EP25, EP33, EP38, EP42, EP43
<i>Text-to-Onto</i>	É um ambiente de aprendizagem de ontologias, baseado em uma arquitetura geral para descobrimento de estruturas conceituais a partir de texto. Ele suporta aprendizado de ontologias a partir de documentos da <i>Web</i> , permite a importação de dados semi-estruturados e estruturados como entrada.	EP3, EP9, EP11, EP13, EP14, EP18, EP19, EP27, EP36, EP38, EP42, EP43
<i>Text2Onto</i>	Text2Onto é uma estrutura de aprendizagem ontologia que foi desenvolvido para apoiar a aquisição de ontologias a partir de documentos textuais. Suportando a geração automática ou semi-automática de ontologias a partir de texto em linguagem natural.	EP7, EP10, EP18, EP20, EP21, EP21, EP22, EP26, EP29, EP31, EP34, EP38, EP40, EP41
<i>OntoLancs</i>	Ferramenta automática de aprendizado de ontologias a partir de texto que fornece suporte para avaliar a utilidade e precisão de diferentes técnicas e as possíveis combinações de técnicas que serão utilizadas para extrair conceitos e transformá-los em uma ontologia.	EP6, EP15, EP39
<i>OntoLT</i>	É uma ferramenta de aprendizado de ontologias de código aberto, com a qual conceitos e relações podem ser extraídas automaticamente a partir de coleções de texto linguisticamente anotados.	EP24, EP42, EP43

Tabela 17 – Ferramentas para aprendizado de ontologias

5.3 Discussão dos Resultados

Apesar do pequeno número de estudos primários que foram selecionados, o mapeamento respondeu de forma satisfatória as questões de pesquisa propostas.

A primeira questão de pesquisa visa descobrir os métodos utilizados no aprendizado de ontologias. Respondendo a esta pergunta, o mapeamento identificou 5 diferentes métodos, conforme descritos na Tabela 15.

A segunda questão de pesquisa pretende descobrir as técnicas utilizadas no aprendizado de ontologia. Respondendo a esta pergunta, o mapeamento identificou 24 técnicas.

As 12 técnicas apresentadas na Tabela 16, sendo que 6 técnicas baseadas em linguísticas, 4 baseadas em estatísticas e 2 baseadas em lógica, foram as mais referenciadas dentre o total de estudos pertencentes ao mapeamento, além das demais técnicas mencionadas.

A terceira questão de pesquisa busca descobrir as ferramentas utilizadas no aprendizado de ontologias. Respondendo a esta questão, foram identificadas 26 ferramentas após a execução do mapeamento. Na Tabela 17 são apresentadas as 8 ferramentas mais referenciadas entre os estudos, além das demais ferramentas mencionadas neste Capítulo.

O gráfico da 15 apresenta a frequência com que essas ferramentas aparecem nos estudos primários.

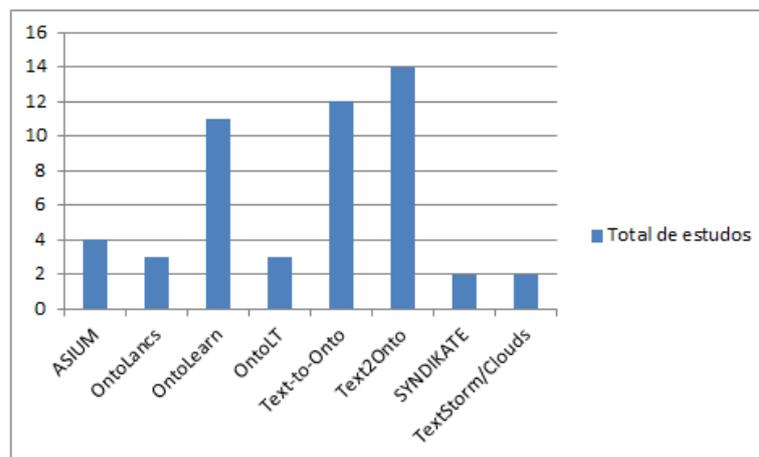


Figura 15 – Frequência de aparecimento das ferramentas nos estudos primários

5.4 Considerações Finais

Neste Capítulo foram apresentados os resultados do mapeamento sistemático realizado. O processo de busca retornou um número total de 1978 estudos, dos quais 43 foram classificados como estudos primários relevantes. *IEEE*, *ACM* e *Scopus* foram as bases que contribuíram para o mapeamento, com respectivamente 29, 5 e 9 estudos primários cada.

Os estudos primários selecionados apontam cinco métodos, vinte-quatro técnicas e vinte-seis ferramentas que são utilizadas no aprendizado de ontologias.

6 Considerações Finais

Neste capítulo, as considerações finais do estudo realizado são apresentadas. Entre elas são discutidas a ameaça à validade do estudo, trabalho futuro e conclusões obtidas com o trabalho.

6.1 Ameaças a Validade

A primeira ameaça a validade diz respeito à estratégia de busca utilizada. A partir do momento em que utilizamos engenhos de buscas automatizados, estudos relevantes podem não ter sido inclusos no conjunto de estudos selecionados. Apesar do tempo dedicado para identificar palavras-chaves relevantes, algum estudo em particular, que usa um termo diferente dos previstos, pode não ser identificado. Além disso, estudos recentes podem ainda não terem sido indexados pelos engenhos de buscas.

Outro fator que pode ser encarado como uma ameaça foi à indisponibilidade de acesso a estudos dos engenhos de busca automatizados e manuais, o que influenciou diretamente na pequena quantidade de estudos primários selecionados, onde apenas 43 estudos foram selecionados ao final.

Por fim, de acordo com (CHARTERS; KITCHENHAM, 2007), o fato dos dados desse trabalho terem sido extraídos por apenas um pesquisador configura uma limitação, mas é um fato aceitado por ela para alunos de PhD (não diz nada a respeito de alunos graduandos).

6.2 Trabalho Futuro

Um importante ponto que deve ser evidenciado em um trabalho é levantar as oportunidades de trabalhos futuros. Visando levantar essas oportunidades, a seguir é apresentado uma sugestão para novos encaminhamentos que foram identificados durante o estudo em questão:

- Realizar experiências utilizando os métodos, técnicas e ferramentas evidenciados a fim de verificar qual se adequam melhor ao desenvolvimento de ontologias para o domínio da Ciência da Computação.

6.3 Conclusão

Este mapeamento sistemático de estudos procura na literatura por estudos apresentaram algum método, técnica ou ferramenta utilizada no aprendizado de ontologias. Três perguntas de pesquisa foram usadas para guiar o estudo e após a avaliação de 1978 estudos, 43 estudos primários foram selecionados.

O processo de execução do mapeamento foi apresentado na monografia permitindo a avaliação e replicação do estudo por terceiros.

A principal contribuição deste estudo esta no resumo dos principais métodos, técnicas e ferramentas utilizados no aprendizado de ontologias, o que facilitara a condução de novos estudos por pesquisadores e desenvolvedores de ontologias em aprendizado de ontologias.

Referências

- AUSABEL, David Paul; NOVAK, Joseph Donald; HANESIAN, Helen. Educational psychology: A cognitive view. 1978. Disponível em: <http://www.projetoderedes.com.br/artigos/artigo_gerenciamento_de_redes_de_computadores.php>. Acesso em: 10 jul 2013.
- BARFOROUSH, Ahmad Abdollahzadeh; SHAMSFARD, Mehrnoush. The state of the art in ontology learning: A framework for comparison. 2004. Disponível em: <<http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.99.5679&rep=rep1&type=pdf>>. Acesso em: 23 out 2012.
- BASÉGIO, Túlio Lima. *Uma abordagem semi-automática para identificação de estruturas ontológicas a partir de textos na língua portuguesa do Brasil*. 124 p. Dissertação (Dissertação) — Pontifícia Universidade Católica do Rio Grande do Sul, Porto Alegre, 2008. Disponível em: <http://tede.pucrs.br/tde_arquivos/4/TDE-2009-06-09T170445Z-1994/Publico/4030.pdf>. Acesso em: 10 jul 2013.
- BERNERS-LEE, Tim; HENDLER, James; LASSILA, Ora. The semantic web. *Scientific American*, 2001. Disponível em: <http://is12918929391.googlecode.com/svn-history/r347/trunk/RPC/Slides/p01_theSemanticWeb.pdf>. Acesso em: 14 dez 2012.
- BIEMANN, Cris. Ontology learning from text: A survey of method. *Journal LDV Forum*, p. 18, 2005. Disponível em: <<http://wortschatz.unileipzig.de/~cbiemann/pub/2005/BiemannLDVOntology05.pdf>>. Acesso em: 14 dez 2012.
- BIOLCHINI, Jorge et al. Systematic review in software engineering. Technical Report ES 679/05, 2005. Disponível em: <<http://www.cin.ufpe.br/~in1037/leitura/systematicReviewSE-COPPE.pdf>>. Acesso em: 16 out 2012.
- BORST, Willen Nico. *Construction of engineering ontologies for knowledge sharing and reuse*. 243 p. Tese (Doutorado) — Center for Telematics and Information Technology, Netherlands, 1997. Disponível em: <<http://www.ub.utwente.nl/webdocs/inf/1/t0000004.pdf>>. Acesso em: 28 out 2012.
- BUITELAAR, Paul; CIMIANO, Philipp; MAGNINI, Bernardo. *Ontology Learning from Text: Methods, Evaluation and Applications*. Amsterdam, The Netherlands: IOS Press, 2005. ISBN 1-58603-523-1.
- BUITELAAR, Paul et al. Linginfo: Design and applications of a model for the integration of linguistic information in ontologies. Proceedings of the OntoLex06 Workshop at LREC, 2006. Disponível em: <<http://www.dfki.de/~romanell/OntoLex2006.pdf>>. Acesso em: 28 nov 2012.

BUITELAAR, Paul; OLEJNIK, Daniel; SINTEK, Michael. A protégé plug-in for ontology extraction from text based on linguistic analysis. Proceedings of the 1st European Semantic Web Symposium (ESWS), Berlin, p. 18, 2004. Disponível em: <<http://olp.dfki.de/OntoLT/esws2004.pdf>>. Acesso em: 28 nov 2012.

CARABALLO, Sharon A. Automatic construction of a hypernym-labeled noun hierarchy from text. Proceedings of the 37th Annual Meeting of the Association for Computational Linguistics (ACL 1999), Association for Computational Linguistics (1999), p. 6, 1999. Disponível em: <<http://acl.ldc.upenn.edu/P/P99/P99-1016.pdf>>. Acesso em: 20 mai 2013.

CHARTERS, Stuart; KITCHENHAM, Barbara A. Guidelines for performing systematic literature reviews in software engineering. EBSE Technical Report, EBSE-2007-01, 2007. Disponível em: <<http://pages.cpsc.ucalgary.ca/~sillito/cpsc-601.23/readings/kitchenham-2007.pdf>>. Acesso em: 17 Out 2012.

CIMIANO, Philipp. *Ontology Learning and Population from Text: Algorithms, Evaluation and Applications*. New York: Springer-Verlag, 2006. 361 p. ISBN 978-0-378-30632-2.

CIMIANO, Philipp; BUITELAAR, Paul; MAGNINI, Bernardo. Ontology learning from text: An overview. Artificial Intelligence and Applications, Amsterdam, The Netherlands, 2005. Disponível em: <[http://ihavhe.com/Doctorado/Articulos_OL/\(Buitelaar2003\)%20OntologyLearningfromTextAnOverview%5B10%5D.pdf](http://ihavhe.com/Doctorado/Articulos_OL/(Buitelaar2003)%20OntologyLearningfromTextAnOverview%5B10%5D.pdf)>. Acesso em: 29 nov 2012.

CIMIANO, Philipp et al. Ontology learning. Steffen Staab; Rudi Studer. (eds.) Handbook on Ontologies. International Handbooks on Information Systems, 2nd edn, pp. 245–267. Springer, Heidelberg, 2009. Disponível em: <<http://hall.org.ua/halls/wizzard/books4/Handbook%20on%20Ontologies%20SE%20-%20S.Staab,%20R.Studer.PDF>>. Acesso em: 19 mar 2013.

COOPER, Harris M. Organizing knowledge syntheses: A taxonomy of literature reviews. Society Knowledge, p. 2, 1988.

FAURE, Claire Nédellec; David. Knowledge acquisition of predicate argument structures from technical texts using machine learning: the system asium. Proceedings of the 11th International Conference on Knowledge Engineering and Knowledge Management (EKAW 1999), Dagstuhl Castle. Lecture Notes in Computer Science, Berlin, p. 5, 1999.

FELLBAUM, Cristiane. Wordnet, an electronic lexical database. MIT Press, 1998. Disponível em: <<http://www.dsoergel.com/cv/C19.pdf>>. Acesso em: 29 nov 2012.

FENSEL, Dieter; DOMINGUE, John; HENDLER, James A. *Handbook of Semantic Web Technologies*. New York: Springer-Verlag Berlin Heidelberg, 2011. 1056 p. ISBN 978-3-540-92912-3.

GAMALLO, Pablo et al. Mapping syntactic dependencies onto semantic relations. Proceedings of the ECAI Workshop on Machine Learning and Natural Language Processing for Ontology Engineering, 2002. Disponível em: <<http://www.sop.inria.fr/acacia/WORKCHOPS/ECAI2002-OLT/Proceedings/Gamallo.pdf>>. Acesso em: 29 nov 2012.

GRUBER, Tom. *What is an ontology?* 1996. Disponível em: <<http://www-ksl.stanford.edu/kst/what-is-an-ontology.html>>. Acesso em: 20 nov 2012>.

GÓMEZ-PÉREZ, Assunción; MANZANO-MACHO, David. A survey of ontology learning methods and techniques. projeto ontoweb. 2003. Disponível em: <<http://www.sti-innsbruck.at/fileadmin/documents/deliverablesOntoweb/D1.5.pdf>>. Acesso em: 29 nov 2012.

GÓMEZ-PÉREZ, Assunción; MANZANO-MACHO, David. An overview of methods and tools for ontology learning from texts. *The Knowledge Engineering Review*, 2005.

HARRIS, Zellig Sabbetai. *Distributional structure*. Word, p. 16, 1954.

HEARST, Marti A. Automatic acquisition of hyponyms from large text corpora. *Proceedings of the 14th International Conference on Computational Linguistics (COLING 1992)*, p. 6, 1992. Disponível em: <<http://acl.ldc.upenn.edu/C/C92/C92-2082.pdf>>. Acesso em: 20 mai 2013.

H.LITTELL, Julia; CORCORAN, Jacqueline; PILLAI, Vijayan. *Systematic reviews and meta-analysis*. Oxford University Press, USA, 2008.

KITCHENHAM, Barbara A. Procedures for performing systematic reviews. Joint Technical Report Software Engineering Group, Keele University (TR/SE-0401), United Kingdom and Empirical Software Engineering, National ICT Australia Ltd, Austrália, 2004. Disponível em: <http://tests-zingarelli.googlecode.com/svn-history/r336/trunk/2-Disciplinas/MethodPesquisa/kitchenham_2004.pdf>. Acesso em: 10 jul 2013.

KITCHENHAM, Barbara A.; DYBÅ, Tore; JØRGENSEN, Magne. Evidence-based software engineering. *ICSE '04: Proceedings of the 26th International Conference on Software Engineering*, Washington, DC, USA, p. 22, 2004. Disponível em: <http://suraj.lums.edu.pk/~cs661s07/02/03%20Current%20State%20of%20Evidence%20Based%20SW%20Eng%20kitchenham04_.pdf>. Acesso em: 10 mar 2013.

KITCHENHAM, Barbara A.; MENDES, Emilia; TRAVASSOS, Guilherme H. A systematic review of cross- vs. within-company cost estimation studies. *Proceedings of EASE 2006*. British Informatics Society Ltd, 2006. Disponível em: <<http://www.cs.auckland.ac.nz/~emilia/publications/EASE2006.pdf>>. Acesso em: 5 mar 2013.

MAEDCHE, Alexander. *Ontology Learning for the Semantic Web*. Massachusetts - USA: Kluwer Academic Publishers, 2002. 241 p. ISBN 0-7293-7656-0.

MAEDCHE, Alexander; STAAB, Steffen. Discovering conceptual relations from text. *Proc. European Conf. Artificial Intelligence (ECAI-00)*, Amsterdam, p. 4, 2000. Disponível em: <<http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.71.1304&rep=rep1&type=pdf>>. Acesso em: 1 dez 2012.

MAEDCHE, Alexander; STAAB, Steffen. *Ontology learning for the semantic web*. *EEE Intelligent Systems, Special Issue on the Semantic Web*, p. 16, 2001. Disponível em: <<http://icc.mpei.ru/documents/00000833.pdf>>. Acesso em: 22 out 2012.

MARCONI, Marina de Andrade; LAKATOS, Eva Maria. *Metodologia Científica*. 6.ed. ed. São Paulo: Atlas, 2007. 315 p.

- NAVIGLI, Roberto; VELARDI, Paola. Learning domain ontologies from document warehouses and dedicated web sites. *Journal Computational Linguistics*, p. 28, 2004. Disponível em: <<http://acl.ldc.upenn.edu/J/J04/J04-2002.pdf>>. Acesso em: 14 dez 2012.
- NOY, Natalya F.; MCGUINNESS, Deborah L. *Ontology development 101: a guide to create your first ontology*. Stanford, p. 25, 2001. Disponível em: <<http://ksl.stanford.edu/people/dlm/papers/ontology-tutorial-noy-mcguinness.doc>>. Acesso em: 20 mai 2013.
- PAI, Madhukar et al. *Systematic reviews and meta-analyses: An illustrated step-by-step guide*. The National Medical Journal of India, India, p. 4, 2004.
- PETERSEN, Kai et al. *Systematic mapping studies in software engineering*. 2th International Conference on Evaluation and Assessment in Software Engineering, 2008. Disponível em: <http://robertfeldt.net/publications/petersen_ease08_sysmap_studies_in_se.pdf>. Acesso em: 19 out 2012.
- PETTCRIEW, Mark; ROBERTS, Helen. *Systematic reviews in the social sciences: A practical guide*. Wiley-Blackwell, 2008. Disponível em: <<http://xa.yimg.com/kg/groups/18751725/462617161/name/manalysis.pdf>>. Acesso em: 29 out 2012.
- RABISER, Rick; GRÜNBACHER, Paul; DHUNGANA, Deepak. *Requirements for product derivation support: Results from a systematic literature review and an expert survey*. *Information and Software Technology*, 2010. Disponível em: <http://ac.els-cdn.com.ez17.periodicos.capes.gov.br/S0950584909001931/1-s2.0-S0950584909001931-main.pdf?_tid=8ba807e2-d889-11e2-b83f-00000aab0f02&acdnat=1371609784_d6f4bdf604895953c6e343731c3cf10f>. Acesso em: 19 jun 2013.
- RANDOLPH, Justus J. *A guide to writing the dissertation literature reviews*. Practical Assessment, Research Evaluation, Walden University, 2009. Disponível em: <<http://pareonline.net/pdf/v14n13.pdf>>. Acesso em: 18 out 2012.
- RUDOLPH, Sebastian. *Acquiring generalized domain range restrictions*. Medina, R., Obiedkov, S. (eds.) *Proceedings of the Sixth International Conference on Formal Concept Analysis (ICFCA 2008)*, Montreal. *Lecture Notes in Artificial Intelligence*, Springer, Heidelberg, p. 13, 2008. Disponível em: <<http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.140.5917&rep=rep1&type=pdf>>. Acesso em: 20 mai 2013.
- RUDOLPH, Sebastian; VOLKER, Johanna. *Fostering web intelligence by semi-automatic owl ontology refinement*. *Proceedings of the Seventh International Conference on Web Intelligence (WI 2008)*, Sydney, Washington, 2008. Disponível em: <Intelligence_by_Semi-automatic_OWL_Ontology_Refinement/file/9fcfd510a8202e0ecb.pdf>. Acesso em: 20 mai 2013.
- SAITO, Andre; UMEMOTO, Katsuhiko; IKEDA, Mitsuru. *strategy-based ontology of knowledge management technology*. *Journal of knowledge Management*, p. 17, 2007. Disponível em: <http://www.ai.wu.ac.at/~kaiser/birgit/Nonaka-Papers-Alfred/SaitoUmamotoIkeda_StrategybasedOntology.pdf>. Acesso em: 02 ago 2012.

SPYNS, Peter; MEERSMAN, Robert; JARRAR, Mustafa. Data modelling versus ontology engineering. *SIGMOD Record*, p. 5, 2002. Disponível em: <<http://www.inf.uni-konstanz.de/dbis/teaching/ws0708/web/papers/MeersmanOntology.pdf>>. Acesso em: 23 nov 2012.

TURNER, Peter D. Mining the web for synonyms: Pmi-ir versus lsa on toefl. *Proceedings of the 12th European Conference on Machine Learning*, p. 11, 2001. Disponível em: <<http://www.inf.uni-konstanz.de/dbis/teaching/ws0708/web/papers/MeersmanOntology.pdf>>. Acesso em: 1 dez 2012.

APÊNDICE A – Calibração das *Strings* de Buscas aplicadas na busca automatizada e resultados da busca manual

A.1 Busca Automática

Base Eletrônica	Quantidade de estudos
<i>IEEEExplore Digital Library</i>	668
<i>ACM Digital Library</i>	346
<i>Elsevier Science Direct</i>	239
<i>Elsevier Scopus</i>	678
Total de Estudos	1931

Tabela 18 – Estudos primários retornados na busca automática

Adaptação da string para cada base eletrônica:

- ***IEEEExplore Digital Library***

(("ontology learning") AND (framework OR method OR methodology OR process OR technique OR techniques) AND (tool OR plugin)) Refinado por: Search: Full Text & Metadata; Publication Year: 2001-2013).

- ***Elsevier Science Direct***

(("ontology learning") AND (framework OR method OR methodology OR process OR technique OR techniques) AND (tool OR plugin)) (Refinado por: Include: Journals; Data Range = 2001-2013; Subject = Computer Science).

- ***ACM Digital Library***

((("ontology learning") AND ("framework"OR "method"OR "methodology"OR "process"OR "technique"OR "techniques") AND ("tool"OR "plugin")) (Refinado por: Full Text & Abstract; Publication Year: 2001-2013).

- *Elsevier Scopus*

OR LIMIT-TO(PUBYEAR, 2002) OR LIMIT-TO(PUBYEAR, 2001)) AND (LIMIT-TO(DOCTYPE, "cp") OR LIMIT-TO(DOCTYPE, "ar")) AND (LIMIT-TO(SUBJAREA, "COMP")) AND (LIMIT-TO(LANGUAGE, "English")) AND (LIMIT-TO(SRCTYPE, "j") OR LIMIT-TO(SRCTYPE, "p"))(("ontology learning") AND (framework OR method OR methodology OR process OR technique OR techniques) AND (tool OR plugin)) AND (LIMIT-TO(PUBYEAR, 2013) OR LIMIT-TO(PUBYEAR, 2012) OR LIMIT-TO(PUBYEAR, 2011) OR LIMIT-TO(PUBYEAR, 2010) OR LIMIT-TO(PUBYEAR, 2009) OR LIMIT-TO(PUBYEAR, 2008) OR LIMIT-TO(PUBYEAR, 2007) OR LIMIT-TO(PUBYEAR, 2006) OR LIMIT-TO(PUBYEAR, 2005) OR LIMIT-TO(PUBYEAR, 2004) OR LIMIT-TO(PUBYEAR, 2003))

OR LIMIT-TO(PUBYEAR, 2002) OR LIMIT-TO(PUBYEAR, 2001)) AND (LIMIT-TO(DOCTYPE, "cp") OR LIMIT-TO(DOCTYPE, "ar")) AND (LIMIT-TO(SUBJAREA, "COMP")) AND (LIMIT-TO(LANGUAGE, "English")) AND (LIMIT-TO(SRCTYPE, "j") OR LIMIT-TO(SRCTYPE, "p")) Refinado por: Limit to Year = 1999-2013; Language = Inglês; Subject Area: Computer Science; Source Type: Conference Proceedings, Journals).

A.2 Busca Manual

Fonte	Estudos Retornados
<i>Semantic Web Journal (SWJ)</i>	24
Journal of Web Semantic (JWS)	23
Total de Estudos	47

Tabela 19 – Estudos primários retornados na busca manual

Descrição sobre a busca:

Na *Semantic Web Journal* foram capturados estudos disponíveis no período entre 2001 a 2013, obtendo acesso aos resultados diretamente do site do *Journal*.

No *Journal of Web Semantic* foram capturados estudos disponíveis no período de 2003 a 2013. Os resultados foram capturados por meio da indexação com a base eletrônica *Elsevier Scienc Direct*.

APÊNDICE B – Estudos Primários

ID	Ano	Fonte	Referência
EP1	2006	IEEE	Klapaftis, I.P. ; Manandhar, S. <i>Term Sense Disambiguation for Ontology Learning</i> . In Intelligent Systems Design and Applications, 2006. ISDA '06. Sixth International Conference on (Volume:2), Page(s): 844 - 849, 2006.
EP2	2010	IEEE	Gil, R. ; Martín-Bautista, M.J. ; Contreras, L. <i>Applying an Ontology Learning Methodology to a Relational Database: University Case Study</i> . In Semantic Computing (ICSC), 2010 IEEE Fourth International Conference on, Page(s): 313 - 316, 2010.
EP3	2011	IEEE	Jaszuk, M. ; Szostek, G. ; Walczak, A. ; Puzio, L. <i>Building a model of disease symptoms using text processing and learning from examples</i> . In Computer Science and Information Systems (FedCSIS), 2011 Federated Conference on, Page(s): 187 - 194, 2011.
EP4	2012	IEEE	Albukhitan, S. ; Helmy, T. <i>Multilingual Food and Health Ontology Learning Using Semi-Structured and Structured Web Data Sources</i> . In Web Intelligence and Intelligent Agent Technology (WI-IAT), 2012 IEEE/WIC/ACM International Conferences on, page(s): 231 - 235, 2012.
EP5	2010	IEEE	Gacitua, R. ; Sawyer, P. ; Gervasi, V. <i>On the Effectiveness of Abstraction Identification in Requirements Engineering</i> . In Requirements Engineering Conference (RE), 2010 18th IEEE International, Page(s): 5 - 14, 2010.
EP6	2009	IEEE	Gacitua, R. ; Arguello Casteleiro, M. ; Sawyer, P. ; Des, J. ; Perez, R. ; Fernandez-Prieto, M.J. ; Paniagua, H. <i>A collaborative workflow for building ontologies: A case study in the biomedical field</i> . In Research Challenges in Information Science, 2009. RCIS 2009. Third International Conference on , Page(s): 121 - 128, 2009.
EP7	2012	IEEE	Toledo, C.M. ; Chiotti, O. ; Galli, M.R. <i>An ontology evolution approach for information retrieval strategies with compound terms</i> . In Informatica (CLEI), 2012 XXXVIII Conferência Latino americana En , Page(s): 1 - 10, 2012.
EP8	2010	IEEE	Sousan, W.L. ; Qiuming Zhu ; Gandhi, R. ; Mahoney, W. ; Sharma, A. <i>Using Term Extraction Patterns to Discover Coherent Relationships from Open Source Intelligence</i> . In Social Computing (SocialCom), 2010 IEEE Second International Conference on, Page(s): 967 - 972, 2010.

EP9	2011	IEEE	Botzenhardt, A. ; Maedche, A. ; Wiesner, J. <i>Developing a domain ontology for software product management</i> . In Software Product Management (IWSPM), 2011 Fifth International Workshop on, Page(s): 7 - 16, 2011. A1:D35
EP10	2010	IEEE	Papageorgiou, N. ; Verginadis, Y. ; Apostolou, D. ; Mentzas, G. <i>Semantic interoperability of e-services in collaborative networked organizations</i> . In e-Business (ICE-B), Proceedings of the 2010 International Conference on, Page(s): 1 - 8, 2010.
EP11	2009	IEEE	Deng Zhonghua ; Liu Shu. <i>Discussion about automatically organizing network education resources with multilingual thesauri-ontology</i> . In IT in Medicine & Education, 2009. ITIME '09. IEEE International Symposium on, Page(s): 326 - 331, 2009.
EP12	2008	IEEE	Medina, M.A. ; Sanchez, J.A. <i>OntOAIr: A Method to Construct Lightweight Ontologies from Document Collections</i> . In Computer Science, 2008. ENC '08. Mexican International Conference on , Page(s): 115 - 125, 2008.
EP13	2007	IEEE	Ming Xiao ; Jinzhu Hu ; Yi Xiao. <i>A Study on Ontology Learning for the Intelligent Search Engine</i> . In Wireless Communications, Networking and Mobile Computing, 2007. WiCom 2007. International Conference on, Page(s): 5369 - 5372, 2007.
EP14	2001	IEEE	Maedche, A. ; Staab, S. <i>Ontology learning for the Semantic Web</i> . In Intelligent Systems, IEEE, Page(s): 72 - 79, 2001.
EP15	2009	IEEE	Arguello, M. ; Gacitua, R. ; Osborne, J. ; Peters, S. ; Ekin, P. ; Sawyer, P. <i>Skeletons and Semantic Web Descriptions to Integrate Parallel Programming into Ontology Learning Frameworks</i> . In Computer Modelling and Simulation, 2009. UKSIM '09. 11th International Conference on, Page(s): 640 - 645, 2009.
EP16	2003	IEEE	Navigli, R. ; Velardi, Paola ; Gangemi, A. <i>Ontology learning and its application to automated terminology translation</i> . In Intelligent Systems, IEEE , Page(s): 22 - 31 , 2003.
EP17	2010	IEEE	Chengxiang Yuan ; Yi Zhuang ; Xiaojun Li. <i>Natural language processing based ontology learning</i> . In Computer Application and System Modeling (ICCASM), 2010 International Conference on, Page(s): V2-700 - V2-704, 2010.
EP18	2011	IEEE	Shubin Cai ; Heng Sun ; Sishan Gu ; Zhong Ming . <i>Learning Concept Hierarchy from Folksonomy</i> . In Web Information Systems and Applications Conference (WISA), 2011 Eighth, Page(s): 47 - 51, 2011.
EP19	2006	IEEE	Karoui, L. ; Afaure, M. ; Bennacer, N. <i>A New Extraction Concept Based on Contextual Clustering</i> . In Computational Intelligence for Modelling, Control and Automation, 2006 and International Conference on Intelligent Agents, Web Technologies and Internet Commerce, International Conference on, Page(s): 91, 2006.

EP20	2011	IEEE	Changuel, S. ; Labroche, N. <i>Distinguishing defined concepts from prerequisite concepts in learning resources.</i> In Computational Intelligence and Data Mining (CIDM), 2011 IEEE Symposium on , Page(s): 22 - 29, 2011.
EP21	2008	IEEE	Mukhlason, A. ; Mahmood, A.K. ; Noreen Izza Arshad,. <i>Semi automatic ontological knowledge base construction from learning materials in eLearning Management System.</i> In Information Technology, 2008. ITSIm 2008. International Symposium on, Page(s): 1 - 8, 2008.
EP22	2011	IEEE	Dai Quoc Nguyen ; Dat Quoc Nguyen ; Khoi Trong Ma ; Pham, S.B. <i>Automatic ontology construction from Vietnamese text.</i> In Natural Language Processing and Knowledge Engineering (NLP-KE), 2011 7th International Conference on, Page(s): 485 - 488, 2011.
EP23	2010	IEEE	Changuel, S. ; Labroche, N. ; Bouchon-Meunier, B. <i>Automatic concept type identification from learning resources.</i> In Neural Networks (IJCNN), The 2010 International Joint Conference on, Page(s): 1 - 6, 2010.
EP24	2008	IEEE	Cerbah, F. <i>Mining the Content of Relational Databases to Learn Ontologies with Deeper Taxonomies.</i> In Web Intelligence and Intelligent Agent Technology, 2008. WI-IAT '08. IEEE/WIC/ACM International Conference on , Page(s): 553 - 557 , 2008.
EP25	2009	IEEE	Siwei Yu ; Hui Dong ; Ying Jiang. <i>Research on the Architecture and the Components of Semantic Web Application Platform.</i> In Information Processing, 2009. APCIP 2009. Asia-Pacific Conference on, Page(s): 605 - 610, 2009.
EP26	2004	IEEE	Cinque, L. ; Malizia, A. ; Navigli, R. <i>OntoDoc: an ontology-based query system for digital libraries.</i> In Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on, Page(s): 671 - 674 Vol.2, 2004.
EP27	2010	IEEE	Chen Shichao ; Yu Bin. <i>Automatic Term Extraction in Cryptology Based on Extension of Central String.</i> In Multimedia Information Networking and Security (MINES), 2010 International Conference on, Page(s): 255 - 258, 2010.
EP28	2008	IEEE	Shih-Hsi Liu ; Yu Cao ; Ming Li ; Kilaru, P. ; Smith, T. ; Toner , S. <i>A Semantics- and Data-Driven SOA for Biomedical Multimedia Systems.</i> In Multimedia, 2008. ISM 2008. Tenth IEEE International Symposium on, Page(s): 533 - 538, 2008.
EP29	2007	IEEE	Suresh, R.M. <i>A study on the ontology based Web mining for digital library.</i> In Information and Communication Technology in Electrical Sciences (ICTES 2007), 2007. ICTES. IET-UK International Conference on , Page(s): 1096 - 1100, 2007.
EP30	2008	IEEE	Feng Yang ; Pei-Guang Lin ; Ru-Zhi Xu. <i>An Emergency-Driven Virtual Organization Model for Emergency Management Based on Ontology.</i> In Wireless Communications, Networking and Mobile Computing, 2008. WiCOM '08. 4th International Conference, Page(s): 1 - 4 , 2008.

EP31	2012	ACM	Wilson Wong, Wei Liu, Mohammed Bennamoun. <i>Ontology learning from text: A look back and into the future</i> . In Computing Surveys (CSUR), Volume 44 Issue 4, August 2012.
EP32	2012	ACM	Angela Locoro, Viviana Mascardi, Anna Marina Scapolla. <i>Evolving and validating annotations in web-based collaborative environments through ontology matching</i> . In SAC '12: Proceedings of the 27th Annual ACM Symposium on Applied Computing, March 2012.
EP33	2011	ACM	Amal Zouaq, Dragan Gasevic, Marek Hatala. <i>Ontologizing concept maps using graph theory</i> . Proceedings of the 2011 ACM Symposium on Applied Computing, March 2011.
EP34	2013	ACM	Csaba Veres, Kristian Johansen, Andreas Opdahl. <i>SynsetTagger: a tool for generating ontologies from semantic tags</i> . In WIMS '13: Proceedings of the 3rd International Conference on Web Intelligence, Mining and Semantics, June 2013.
EP35	2008	ACM	Michela Pedroni, Manuel Oriol, Bertrand Meyer, Lukas Angerer. <i>Automatic extraction of notions from course material</i> . In SIGCSE '08: Proceedings of the 39th SIGCSE technical symposium on Computer science education, February 2008.
EP36	2012	Scopus	Jung, J.J. <i>Evolutionary approach for semantic-based query sampling in large-scale information sources</i> . In Information Sciences 182 (1), pp. 30-39, 2012.
EP37	2011	Scopus	Meij, E., Bron, M., Hollink, L., Huurnink, B., De Rijke, M. <i>Mapping queries to the Linking Open Data cloud: A case study using Dbpedia</i> . In Journal of Web Semantics 9 (4), pp. 418-433, 2011.
EP38	2011	Scopus	Frikh, B., Djaanfar, A.S., Ouhbi, B. <i>A new methodology for domain ontology construction from the web</i> . In International Journal on Artificial Intelligence Tools 20 (6), pp. 1157-1170, 2011.
EP39	2011	Scopus	Zouaq, A., Gasevic, D., Hatala, M. <i>Towards open ontology learning and filtering</i> . In Information Systems 36 (7), pp. 1064-1081, 2011.
EP40	2011	Scopus	Hou, X., Ong, S.K., Nee, A.Y.C., Zhang, X.T., Liu, W.J. <i>GRAONTO: A graph-based approach for automatic construction of domain ontology</i> . In Expert Systems with Applications 38 (9), pp. 11958-11975, 2011.
EP41	2009	Scopus	Ottens, K., Hernandez, N., Gleizes, M.-P., Aussenac-Gilles, N. <i>A multi-agent system for dynamic ontologies</i> . In Journal of Logic and Computation 19 (5), pp. 831-858, 2009.
EP42	2008	Scopus	McDowell, L.K., Cafarella, M. <i>Ontology-driven, unsupervised instance population</i> . In Web Semantics 6 (3), pp. 218-236, 2008.
EP43	2007	Scopus	Jing, K. <i>Ontology learning for Chinese information organization and knowledge discovery in ethnology and anthropology</i> . In Data Science Journal 6 (SUPPL.), pp. S500-S510, 2007.
EP44	2005	Scopus	Sabou, M., Wroe, C., Goble, C., Stuckenschmidt, H. <i>Learning domain ontologies for semantic Web service descriptions</i> . In Web Semantics 3 (4), pp. 340-365, 2005.