

UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ
CÂMPUS DE CORNÉLIO PROCÓPIO
DIRETORIA DE PESQUISA E PÓS-GRADUAÇÃO PROGRAMA DE
PÓS-GRADUAÇÃO EM INFORMÁTICA

VINICIUS DOS SANTOS

**CONSTRUÇÃO AUTOMÁTICA DE RESUMOS GRÁFICOS
UTILIZANDO PROCESSAMENTO DE LINGUAGEM NATURAL**

DISSERTAÇÃO DE MESTRADO

CORNÉLIO PROCÓPIO

2018

VINICIUS DOS SANTOS

**CONSTRUÇÃO AUTOMÁTICA DE RESUMOS GRÁFICOS
UTILIZANDO PROCESSAMENTO DE LINGUAGEM NATURAL**

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Informática da Universidade Tecnológica Federal do Paraná como requisito para obtenção do título de “Mestre profissional em Informática”.

Orientador: Prof^a Dr^a Érica Ferreira Souza

Co-orientador: Prof^a Dr^a Katia Romero Felizardo
Scannavino

CORNÉLIO PROCÓPIO

2018



Título da Dissertação Nº 45:

“CONSTRUÇÃO AUTOMÁTICA DE RESUMOS GRÁFICOS UTILIZANDO PROCESSAMENTO DE LINGUAGEM NATURAL”.

por

Vinicius dos Santos

Orientador: **Profa. Dra Érica Ferreira de Souza**

Esta dissertação foi apresentada como requisito parcial à obtenção do grau de MESTRE EM INFORMÁTICA – Área de Concentração: Computação Aplicada, pelo Programa de Pós-Graduação em Informática – PPGI – da Universidade Tecnológica Federal do Paraná – UTFPR – Câmpus Cornélio Procópio, às 14h do dia 02 de abril de 2018. O trabalho foi _____ pela Banca Examinadora, composta pelos professores:

Profa. Dra. Érica Ferreira de Souza
(Presidente – UTFPR-CP)

Profa. Dra. Katia Romero Felizardo Scannavino
(Co-orientadora - UTFPR-CP)

Prof. Dr. Willian Massami Watanabe
(UTFPR-CP)

Prof. Dr. Nandamudi Lankalapalli Vijaykumar
(LAC/INPE)
Participação à distância via _____

Prof. Dr. Arnaldo Cândido Júnior
(UTFPR-MD)

Participação à distância via _____

Visto da coordenação:

Danilo Sipoli Sanches
Coordenador do Programa de Pós-Graduação em Informática
UTFPR Câmpus Cornélio Procópio

A Folha de Aprovação assinada encontra-se na Coordenação do Programa.

Dedico este trabalho à minha família.

AGRADECIMENTOS

Gostaria de agradecer primeiramente à Deus, pois sem sua força não teria nada e todos os meus dias são dedicados a Ele. Agradeço ainda a minha família, que me apoia em todos os momentos e me dá o suporte necessário para que eu possa seguir em frente.

Agradeço a todos os professores que participaram direta e indiretamente do processo de elaboração deste projeto. De forma especial gostaria agradecer ao professor Dr. Nandamudi L. Vijaykumar que forneceu ideias muito importantes e participou diretamente da condução do projeto.

Ainda devo um agradecimento mais que especial ao professor Dr. **Willian Massami Watanabe** que me acompanhou durante todo o mestrado e sempre me passou segurança e motivação para seguir em frente. A presença do professor Willian foi essencial para realização deste trabalho.

Por fim, gostaria de ressaltar a enorme gratidão à minha orientadora Dr^a **Érica Ferreira de Souza** e coorientadora Dr^a **Katia Romero Felizardo Scannavino** que acreditaram em mim desde o primeiro contato e fizeram de tudo para me apoiar nesse passo tão importante da minha vida acadêmica.

Se as pessoas não acreditam que a matemática é simples, é só porque eles não percebem o quanto complicada a vida é.

John Von Neumann

RESUMO

SANTOS, Vinicius dos. CONSTRUÇÃO AUTOMÁTICA DE RESUMOS GRÁFICOS UTILIZANDO PROCESSAMENTO DE LINGUAGEM NATURAL. 116 f. DISSERTAÇÃO DE MESTRADO – Diretoria de pesquisa e pós-graduação programa de pós-graduação em informática, Universidade Tecnológica Federal do Paraná . Cornélio Procópio, 2018.

Contexto: Estudos secundários, tais como Revisões Sistemáticas da Literatura (RSL) e Mapeamentos Sistemáticos (MS), têm sido cada vez mais utilizados na Engenharia de Software (ES) uma vez que permitem a identificação de evidências disponíveis relacionadas com um tópico de pesquisa. Uma das principais atividades do processo de condução de um estudo secundário é a seleção dos estudos primários, que envolve, em um primeiro momento, a leitura dos resumos dos estudos candidatos. No entanto, com o crescente número de publicações científicas, agregado à baixa qualidade dos seus resumos, torna essa atividade cada vez mais difícil para os pesquisadores. Algumas soluções têm sido propostas para atenuar o problema, entre elas, a utilização dos resumos gráficos baseados em Mapas Conceituais (MC). No entanto, estes resumos são criados de forma manual. **Objetivo:** este trabalho possui dois objetivos: (i) entender o uso dos MCs na Ciência da Computação, bem como identificar as principais técnicas para geração de MCs a partir do Processamento de Linguagem Natural (PLN); e (ii) propor uma abordagem para a construção automática de resumos gráficos baseados em MCs por meio de técnicas de PLN. **Método:** inicialmente foram conduzidos dois mapeamentos sistemáticos da literatura para entender o uso dos MCs na ciência da computação e identificar as principais práticas para construção de MCs a partir de PLN. Em seguida, foi definida uma abordagem para construção de resumos gráficos baseados em MCs. Por fim, foram conduzidas avaliações com o objetivo de verificar a qualidade dos MCs gerados. **Resultado:** O experimento piloto realizado apresentou que os MCs construídos pela iniciativa demonstraram um bom desempenho em termos de extração de conceitos e abrangência ao representar o conteúdo do resumo. **Conclusão:** Os resultados preliminares apresentados demonstram que a iniciativa proposta pode gerar proposições válidas e representar resumos gráficos por meio de MCs, tornando-se um importante instrumento para sumarizar uma estrutura complexa de informações textuais, contribuindo para a identificação das informações mais importantes de um artigo.

Palavras-chave: Mapas Conceituais, Processamento de Linguagem Natural, Revisões Sistemáticas, Mapeamentos Sistemáticos

ABSTRACT

SANTOS, Vinicius dos. AUTOMATIC CONSTRUCTION OF GRAPHIC SUMMARIES USING NATURAL LANGUAGE PROCESSING. 116 f. DISSERTAÇÃO DE MESTRADO – Diretoria de pesquisa e pós-graduação programa de pós-graduação em informática, Universidade Tecnológica Federal do Paraná . Cornélio Procópio, 2018.

Context: Secondary studies, such as Systematic Literature Reviews (SLR) and Systematic Mappings (SM), have been increasingly used in Software Engineering (SE) since they allow the identification of available evidence related to a research topic. One of the main activities of the process of conducting a secondary study is the primary studies selection, which involves, at first, the reading of the abstracts of the candidate studies. However, with the growing number of scientific publications, coupled with the poor quality of their abstracts, it makes this activity increasingly difficult for researchers. Some solutions have been proposed to mitigate the problem, among them, the use of structured abstracts and graphic summaries. Previous studies have proposed guidelines for the construction of graphic summaries. However, these summaries continue to be created manually. **Objectives:** This work has two objectives: (i) understand the use of Conceptual Maps (CM) in Computer Science and to investigate the main techniques for generation of MCs from Natural Language Processing (NLP); (ii) propose an approach for the automatic construction of graphic abstracts based on CMs using NLP techniques. **Method:** initially the collection of the main practices for the construction of CMs from NLP was performed. Next, an approach for the construction of graphic summaries based on CMs was defined. Finally, evaluations were conducted in order to verify the quality of the CMs generated. **Results:** The pilot experiment conducted showed that the CMs constructed by the initiative demonstrated a good performance in terms of concept extraction and comprehensiveness when representing the concepts of the abstract. **Conclusions:** The preliminary results show that the proposed initiative can generate valid propositions and represent graphic summaries through CMs, becoming an important tool to summarize a complex structure of textual information, contributing to the identification of the most important information of an article.

Keywords: Concept Map, Natural Language Processing, Systematic Review, Systematic Mapping

LISTA DE FIGURAS

FIGURA 1	– Processo de condução de uma RS	17
FIGURA 2	– Exemplo de um MC	19
FIGURA 3	– Estrutura proposta para representação de um resumo gráfico	21
FIGURA 4	– Fluxograma do MS - MCs na Ciência da Computação	27
FIGURA 5	– Mapa de conteúdo gerado a partir dos estudos selecionados.	29
FIGURA 5	– Processo de seleção conduzido no MS	38
FIGURA 6	– Distribuição dos estudos selecionados ao longo dos anos	41
FIGURA 7	– Propósito de aplicação dos estudos	42
FIGURA 8	– Nível de automatização dos estudos	42
FIGURA 9	– Técnicas utilizadas para construção de MCs	43
FIGURA 10	– Ferramentas para construção de MCs	45
FIGURA 11	– Algoritmos para construção de MCs	47
FIGURA 12	– <i>Framework</i> para construção dos MCs baseados em PLN	48
FIGURA 13	– Grafo de soluções	49
FIGURA 14	– Iniciativa de extração e sumarização de MCs	52
FIGURA 15	– Árvore de decisão gerada pelo classificador que não utiliza <i>stemmer</i> ...	57
FIGURA 16	– Avaliação do classificador de resumos considerando a classe qualitativa ..	58
FIGURA 17	– Avaliação do classificador de resumos considerando a classe quantitativa ..	58
FIGURA 18	– Exemplo de saída do algoritmo de <i>POS-Tagging</i>	61
FIGURA 19	– Exemplo criado pela abordagem proposta	62
FIGURA 20	– Avaliação dos conceitos apresentados nos MCs gerados automaticamente ..	66
FIGURA 21	– Avaliação dos conceitos apresentados nos MCs gerados pelos autores ..	66
FIGURA 22	– Avaliação dos relacionamento apresentados nos MCs gerados automaticamente ..	67
FIGURA 23	– Avaliação dos relacionamento apresentados nos MCs gerados pelos autores ..	68
FIGURA 24	– Avaliação da abrangência dos MCs gerados automaticamente	68
FIGURA 25	– Avaliação da abrangência dos MCs gerados pelos autores	69

LISTA DE TABELAS

TABELA 1	– String de busca - Aplicação dos MCs na Ciência da Computação	26
TABELA 2	– Checklist da avaliação de qualidade	29
TABELA 3	– String de busca - Construção de MCs utilizando PLN	37
TABELA 4	– Anais da Conferência Internacional de MCs usado na busca manual ...	39
TABELA 5	– Lista de IDs e referências	39
TABELA 6	– Classificações e estudos encontrados pelas QP2, QP3 e QP5	44
TABELA 7	– Relação de algoritmos encontrados em cada estudo	46

LISTA DE SIGLAS

ESBE	Engenharia de Software Baseada em Evidências
RSL	Revisão Sistemática de Literatura
MS	Mapeamento Sistemático
MC	Mapas Conceituais
PLN	Processamento de Linguagem Natural
RS	Revisão Sistemática
ES	Engenharia de Software
IA	Inteligência Artificial
CLN	Compreensão de Linguagem Natural
GLN	Geração de Linguagem Natural
POO	Programação Orientada a Objetos
MMC	Mineração de Mapas Conceituais
XML	<i>Extensible Markup Language</i>
IE	<i>Information Extractor</i>
LSA	<i>Latent Semantic Analysis</i>
PCA	<i>Principal Component Analysis</i>

SUMÁRIO

1 INTRODUÇÃO	12
1.1 MOTIVAÇÃO	13
1.2 OBJETIVOS	15
1.3 ORGANIZAÇÃO DO TEXTO	15
2 FUNDAMENTAÇÃO TEÓRICA	16
2.1 ESTUDOS SECUNDÁRIOS	16
2.2 MAPAS CONCEITUAIS	18
2.3 PROCESSAMENTO DE LINGUAGEM NATURAL	22
3 MAPAS CONCEITUAIS APLICADOS NA CIÊNCIA DA COMPUTAÇÃO	25
3.1 PROTOCOLO DA REVISÃO	25
3.2 SELEÇÃO DOS ESTUDOS	27
3.3 RESULTADOS DO MAPEAMENTO	29
3.4 CONSIDERAÇÕES FINAIS	34
4 CONSTRUÇÃO DE MAPAS CONCEITUAIS A PARTIR DE PLN	36
4.1 PROTOCOLO DE REVISÃO	36
4.2 SELEÇÃO DE ESTUDOS	37
4.2.1 Classificação dos Estudos Selecionados	40
4.3 RESULTADOS DO MAPEAMENTO	41
4.3.1 <i>Framework</i> para criação de MCS baseado em PLN	47
4.4 CONSIDERAÇÕES FINAIS	50
5 ABORDAGEM PARA GERAÇÃO DE MCS A PARTIR DE PLN	51
5.1 ENTRADA - RESUMOS ESTRUTURADOS	51
5.2 PIPELINE 1 - EXTRAÇÃO DE CONCEITOS E RELACIONAMENTOS	53
5.3 PIPELINE 2 - CLASSIFICAÇÃO	54
5.4 SAÍDA - SUMARIZAÇÃO DOS MCS	54
5.5 CONSIDERAÇÕES FINAIS	55
6 VALIDAÇÃO DA ABORDAGEM	56
6.1 AVALIAÇÃO DA EFICIÊNCIA DO CLASSIFICADOR	56
6.2 VALIDAÇÃO DOS MCS CONSTRUÍDOS AUTOMATICAMENTE	59
6.2.1 Construindo um MC automaticamente	59
6.2.2 Condução do estudo experimental	62
6.2.3 Resultados do experimento	65
6.3 DISCUSSÃO	67
7 CONCLUSÃO	71
7.1 PRINCIPAIS CONTRIBUIÇÕES E LIMITAÇÕES DA PESQUISA	71
7.2 TRABALHOS FUTUROS	71
7.3 PUBLICAÇÕES	72
REFERÊNCIAS	73
Apêndice A – TREINAMENTO	80
Apêndice B – EXECUÇÃO - MCS GERADOS AUTOMATICAMENTE	85
Apêndice C – EXECUÇÃO - MCS GERADOS PELOS AUTORES	101

1 INTRODUÇÃO

A Engenharia de Software Baseada em Evidências (ESBE) foi proposta em 2004 por Kitchenham et al. (2004). A ESBE tem como objetivo prover meios para que as melhores evidências de pesquisas sejam integradas com a experiência prática e valores humanos nas tomadas de decisão sobre o desenvolvimento e manutenção de software. O principal objetivo da ESBE é aproximar a pesquisa da prática na Engenharia de Software (ES) disponibilizando meios para construir um corpo de conhecimento sobre quando, como e quais processos, tecnologias, ferramentas, entre outros, são adequados para serem utilizados na ES.

Dentre os métodos da ESBE, destacam-se as Revisões Sistemáticas da Literatura (RSL) e os Mapeamentos Sistemáticos (MS). De acordo com Kitchenham e Charters (2007), existem dois tipos de estudos complementares, aqueles que são denominados primários (estudos experimentais, estudos de caso ou *surveys*) e os denominados secundários. As RSLs e os MSs são considerados estudos secundários.

Estudos secundários têm sido cada vez mais utilizados na ES uma vez que permitem a identificação de evidências disponíveis relacionadas com um particular tópico de pesquisa. Uma RSL tem como objetivo reunir dados de diferentes estudos primários para sintetizar o conhecimento sobre um tópico de interesse e responder uma determinada questão de pesquisa. Da mesma forma que uma RSL, o MS tem como objetivo identificar e classificar os estudos primários, porém é considerado uma revisão mais ampla das evidências disponíveis (KITCHENHAM; CHARTERS, 2007).

De forma geral, o processo para condução de um estudo secundário envolve três principais fases (KITCHENHAM, 2004): *o planejamento da revisão*, a *condução da revisão* e a *divulgação dos resultados*. Na fase de planejamento, um protocolo para condução da revisão é criado. No protocolo são especificados os procedimentos a serem seguidos para conduzir um estudo secundário, tais como questões de pesquisa a serem respondidas, definição das bases de busca, critérios de seleção (inclusão e exclusão) e critérios de qualidade dos estudos selecionados. É durante a etapa de condução da revisão que estudos são identificados e selecionados por

meio da aplicação de critérios de seleção. Após a atividade de seleção, os dados contidos nos estudos incluídos devem ser extraídos e sintetizados. Por fim, na última etapa, os resultados das respostas das questões de pesquisas devem ser divulgados aos potenciais interessados.

Dentre atividades do processo de condução de um estudo secundário, uma das principais atividades é a seleção dos estudos primários. A atividade de seleção é conduzida em três fases: seleção inicial, seleção final e revisão da seleção. Durante a fase de seleção inicial, os títulos e os resumos dos estudos candidatos são lidos e classificados como incluídos ou excluídos por meio da aplicação dos critérios de seleção. Na seleção final os estudos incluídos na fase anterior são lidos na íntegra e novamente os critérios de seleção são aplicados. Durante a revisão da seleção uma amostra dos estudos é revisada para garantir que os critérios de seleção foram aplicados de maneira correta.

1.1 MOTIVAÇÃO

Os estudos secundários possuem métodos bem definidos para realização do planejamento, condução e documentação dos resultados. Essa característica traz benefícios como a possibilidade de replicação da pesquisa e a realização de auditoria no processo de busca. Apesar disso, a etapa de seleção acaba se tornando complexa, principalmente quando os revisores precisam trabalhar com um grande volume de dados. Se os resumos dos estudos primários forem mal estruturados ou escritos podem atrapalhar os pesquisadores em sua busca (BRERETON et al., 2007). Dessa forma, considera-se necessário dar suporte a elaboração de resumos e consequentemente garantir sua qualidade.

Diversos estudos têm investigado o potencial benefício dos resumos estruturados como uma possível solução para melhorar a qualidade dos resumos (KITCHENHAM et al., 2008; BUDGEN et al., 2011). Atualmente, existem diretrizes para a criação de resumos estruturados na ES (BUDGEN et al., 2011). Tais resumos baseiam-se na utilização de um conjunto de seções, como: contexto, objetivo, método, resultados e conclusão (KITCHENHAM et al., 2008). Essas seções têm como objetivo orientar os autores durante a escrita do resumo. Uma segunda alternativa para promover a melhoria dos resumos é a utilização de resumos gráficos. As representações visuais podem ser uma boa opção para os resumos tradicionais, uma vez que permitem a manipulação e interação de uma grande quantidade de dados sumarizando o conteúdo do artigo de forma concisa por meio de uma imagem (CRUZES; DYBÅ, 2010).

De acordo com Jarvenpaa e Dickson (1988), as representações visuais são as mais indicadas para resumir dados. Os Mapas Conceituais (MC) (NOVAK; CAÑAS, 2008), por

exemplo, são ferramentas úteis para representar o conhecimento e recuperar informações. Um MC permite a estruturação e a síntese do conhecimento, além de possibilitar a visualização de elementos relacionados facilitando a extração de conhecimento (NOVAK; CAÑAS, 2008). Dessa forma, o uso de MC pode ser uma alternativa para representar resumos gráficos e podem ser um instrumento importante para sumarizar uma estrutura complexa de informações textuais, contribuindo para a identificação das informações mais importantes de um artigo. No entanto, de acordo com Santos et al. (2017), na construção de MCs alguns problemas podem surgir:

- Dificuldade em criar os MCs: é importante ter o conhecimento da área para desenvolver um MC de qualidade. Caso o autor não domine a área ele pode ter problemas para encontrar os conceitos e relacionamentos corretos.
- Sobrecarga cognitiva: os MCs podem apresentar uma fraqueza; eles podem conter muitos conceitos em sua estrutura causando sobrecarga cognitiva, tornando difícil entender o conteúdo apresentado.
- Escalabilidade: MCs com um grande número de conceitos e relacionamentos podem se tornar muito complexos. Um MC pode perder sua utilidade quando a quantidade de informações aumenta.

A construção automática de MCs pode solucionar ou pelo menos minimizar tais problemas. O Processamento de Linguagem Natural (PLN), por exemplo, pode ser utilizado para extraír conceitos e relacionamentos de textos de forma automática (QASIM et al., 2013). O PLN é uma subárea da Inteligência Artificial que têm foco na extração de informações de bases de dados não estruturadas. O PLN é compreendido como uma gama de técnicas computacionais para análise automática e representação da linguagem humana (ARANHA, 2007). A pesquisa em PLN têm focado em atividades como extração de informações, processamento de texto, modelagem de tópicos e mineração de opiniões. Os sistemas de PLN capturam o significado das palavras inseridas (sentenças, parágrafos, páginas) em forma de um resultado estruturado que pode variar dependendo da aplicação.

Dado o contexto acima, Takemiya e Felizardo (2016) buscam incentivar a utilização de resumos gráficos para auxiliar a etapa de seleção de estudos secundários na ES. Para isso é proposta a elaboração de um modelo de resumo gráfico utilizando MCs a partir de conjunto de diretrizes para a instanciação desse modelo. Como continuação do projeto desenvolvido por Takemiya e Felizardo (2016), este trabalho busca apoiar o processo de seleção de estudos primários a partir da construção dos MCs de forma automática aplicando técnicas de PLN.

1.2 OBJETIVOS

Este trabalho possui dois objetivos: (i) compreender o uso dos MCs na Ciência da Computação e identificar as principais técnicas para geração de MCs a partir do PLN; e (ii) definir uma abordagem para a construção automática de resumos gráficos baseados em MCs por meio de técnicas de PLN.

Para alcançar o primeiro objetivo proposto as seguintes atividades foram conduzidas:

- Condução de um MS para identificar o estado da arte sobre o uso dos MCs na área da Ciência da Computação.
- Condução de um MS para sumarizar as principais iniciativas de construção de MCs a partir de técnicas de PLN.

Considerando os resultados alcançados pelos MSs conduzidos, foram definidas as atividades para alcançar o segundo objetivo deste projeto de mestrado. As seguintes atividades foram conduzidas:

- Definição de uma abordagem para geração automática de MCs com base em técnicas de PLN. Para essa etapa foram consideradas as diretrizes propostas por Takemiya e Felizardo (2016) para representação de resumos gráficos a partir de MCs.
- Condução de avaliações com o objetivo de verificar a qualidade de MCs gerados pela abordagem proposta.

1.3 ORGANIZAÇÃO DO TEXTO

Nesse capítulo foi apresentado o contexto no qual este trabalho se insere, a motivação para a sua realização e os objetivos a serem alcançados. No Capítulo 2, são apresentados conceitos sobre estudos secundários, MCs e PLN. No Capítulo 3, são apresentados os resultados do MS que buscou mapear como os MCs são utilizados na Ciência da Computação. No Capítulo 4, são apresentados os resultados do MS que buscou identificar iniciativas que buscam construir MCs a partir de técnicas de PLN. No Capítulo 5, é apresentada a abordagem proposta e seu funcionamento, sendo descritas as principais atividades realizadas. O Capítulo 6, apresenta como foram feitas as validações da abordagem proposta, os resultados obtidos com a aplicação da abordagem e discussão dos resultados. Por fim, no Capítulo 7, são apresentadas as considerações finais e trabalhos futuros.

2 FUNDAMENTAÇÃO TEÓRICA

Nesse capítulo são descritos os principais conceitos relacionados ao projeto. Na Seção 2.1 é apresentada uma visão geral sobre os estudos secundários. Na Seção 2.2 são apresentadas as características dos MCs. Por fim, na Seção 2.3 são apresentadas as principais definições de PLN.

2.1 ESTUDOS SECUNDÁRIOS

As Revisões Sistemáticas da Literatura (RSL), ou simplesmente Revisões Sistemáticas (RS), foram originalmente instituídas na área da medicina. Porém, com o crescente número de estudos publicados área, criou-se uma necessidade de que os profissionais constantemente se atualizassem do que estava sendo praticado e descoberto por pesquisadores.

Assim como na medicina, a Ciência da Computação também possui constante evolução. Com a popularização da técnica de RS, Kitchenham (2004) propôs adaptar este método para a Engenharia de Software (ES). O processo proposto foi dividido em 3 principais etapas: *Planejamento da Revisão, Condução da Revisão e Divulgação dos Resultados*. A Figura 1 apresenta quais são as atividades que compõem cada uma dessas etapas.

Na etapa de planejamento do processo de revisão é identificada a motivação para a condução de uma RS. Uma vez definidas as necessidades para conduzir uma RS, é definido um protocolo. Todos procedimentos para condução da RS são definidos nesse protocolo. Esse documento contém os detalhes da condução do estudo possibilitando sua replicação posterior por outros pesquisadores. São definidos no protocolo a descrição dos objetivos, as questões de pesquisa a serem respondidas, a estratégia de busca, a *string* de busca, as fontes de pesquisa, os critérios de inclusão e exclusão, dentre outros.

Após a definição do protocolo, é necessário que se avalie sua qualidade. Nesta atividade Kitchenham (2004) propõe que sejam testadas as *strings* de busca e validando se os estudos relevantes para a pesquisa estão sendo retornados. Pode-se utilizar um grupo de controle que

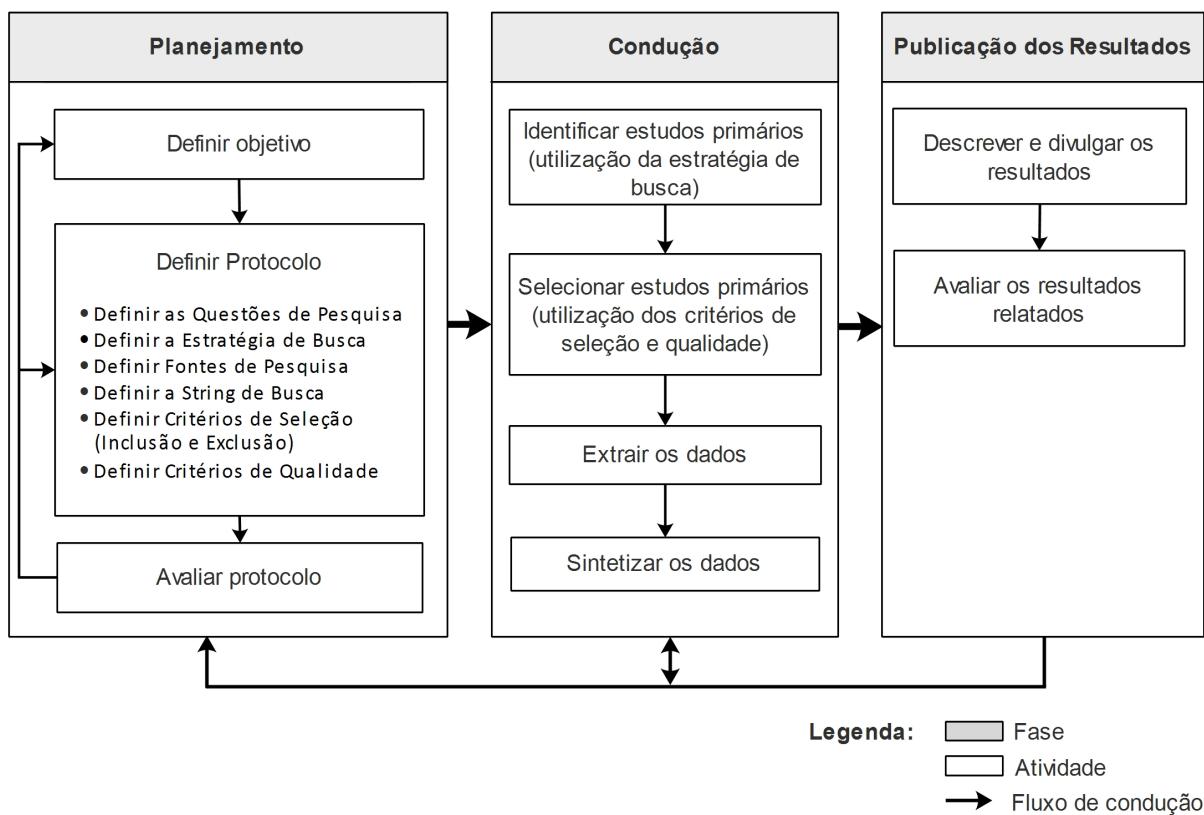


Figura 1: Processo de condução de uma RS

Fonte: (NAKAGAWA et al., 2017)

contém estudos relevantes para a pesquisa. De acordo com Felizardo et al. (2016), um grupo de controle é conjunto de estudos que devem ser retornados a partir da condução da revisão.

Na segunda etapa da revisão, na execução da revisão, o objetivo é encontrar estudos primários capazes de responder as questões de pesquisa. Esta busca pode acontecer de forma manual ou automática. A busca manual ocorre na visita dos anais dos eventos ou revistas científicas sobre o domínio e busca estudos que tratam um determinado assunto. A busca automática acontece na consulta de bases de dados eletrônicas (*IEEE Xplore, ACM Digital Library, Springer, Science Direct*) por meio da *string* de busca definida no protocolo.

A partir dos estudos coletados nas bases de dados, o pesquisador realiza a seleção dos estudos a partir da aplicação dos critérios de inclusão e exclusão. Para otimização deste processo, são adotadas algumas etapas para a atividade de seleção dos dados. Primeiramente, é realizada uma seleção inicial em que se aplicam os critérios de inclusão e exclusão apenas no título, resumo e palavras-chave. Caso o pesquisador identifique que o estudo não se enquadra dentro dos critérios de busca, ele é excluído. Em seguida é conduzida uma seleção final. Na seleção final os estudos incluídos na atividade anterior são lidos na íntegra e novamente os critérios de inclusão e exclusão são aplicados. Ao final do processo de seleção, uma amostra

dos estudos é revisada para garantir que os critérios de inclusão e exclusão foram aplicados de maneira correta.

Após a atividade de seleção, os dados contidos nos estudos primários devem ser extraídos e sumarizados. Formulários de extração de dados são utilizados para coletar os dados que sejam necessários para responder às questões de pesquisa da revisão e para facilitar posteriormente as análises e síntese dos resultados.

A última etapa está relacionada a escrita e divulgação dos resultados para os potenciais interessados. Isto pode ser realizado por meio da publicação de relatórios técnicos, artigos de revistas ou conferências, capítulos de livros e trabalhos de conclusão de curso (dissertações de mestrado ou teses de doutorado).

Da mesma forma que as RSs, os Mapeamentos Sistemáticos (MSs) também são estudos secundários e, nesse caso, eles têm como objetivo identificar e classificar os estudos primários relacionados com um tópico de pesquisa, porém são considerados uma revisão mais ampla das evidências disponíveis. Apesar de ambos serem considerados estudos secundários e terem os processo de condução muito similar, um MS é indicado quando não se têm uma visão definida sobre a área de pesquisa e se deseja apenas quantificar as evidências existentes. Na RS o pesquisador já conhece a área de pesquisa e o objetivo é aprofundar o estudo em um tópico específico. O MS possui questões de pesquisas mais genéricas que buscam explorar a área e compreender o domínio, enquanto uma RS possui questões de pesquisa mais específicas que buscam esclarecer um ponto específico dentro do domínio (PETERSEN et al., 2015).

Mesmo apresentando benefícios, a execução de um estudo secundário pode ser tediosa e morosa, principalmente na atividade de seleção de estudos. A seguir serão apresentadas as características que motivaram a escolha dos MCs como representação visual para criação dos resumos gráficos em artigos científicos.

2.2 MAPAS CONCEITUAIS

O MC foi desenvolvido em 1972, dentro do programa de pesquisa realizado pelo professor Joseph Donald Novak, na Universidade de Cornell (NOVAK; MUSONDA, 1991; NOVAK; CAÑAS, 2006). Os MCs surgiram da necessidade de encontrar a melhor maneira de representar a compreensão conceitual das crianças relacionadas a área de ciências. Ao criar os MCs, Novak propôs que esta representação gráfica fosse capaz de expressar de forma clara as mudanças nos conceitos de estruturas (NOVAK; CAÑAS, 2006). O programa de pesquisa de Novak baseava-se na psicologia da aprendizagem de David Ausubel (AUSUBEL, 1963; AUSU-

BEL et al., 1968, 1978). O autor argumenta que a aprendizagem ocorre por meio da assimilação de novos conceitos e proposições, porém esses conceitos devem ser conectados com outros sistemas proposicionais já adquiridos pelo aprendiz (NOVAK, 2003).

Segundo Novak (2003), os MCs são ferramentas gráficas para a organização e representação do conhecimento. Eles incluem conceitos, geralmente dentro de círculos ou quadros, e relações entre eles, que são indicadas por linhas. Nessas linhas há palavras ou frases de ligação, que especificam os relacionamentos entre dois conceitos. Um exemplo de um MC está representado na Figura 2 que apresenta a estrutura dos MCs e suas características. Uma característica importante dos MCs está relacionada a estrutura na qual os conceitos estão dispostos no mapa. O MC pode seguir um modelo hierárquico no qual os conceitos mais inclusivos estão no topo da hierarquia, ou seja, na parte superior do mapa. Já os conceitos mais específicos, pouco abrangentes estão na base da hierarquia.

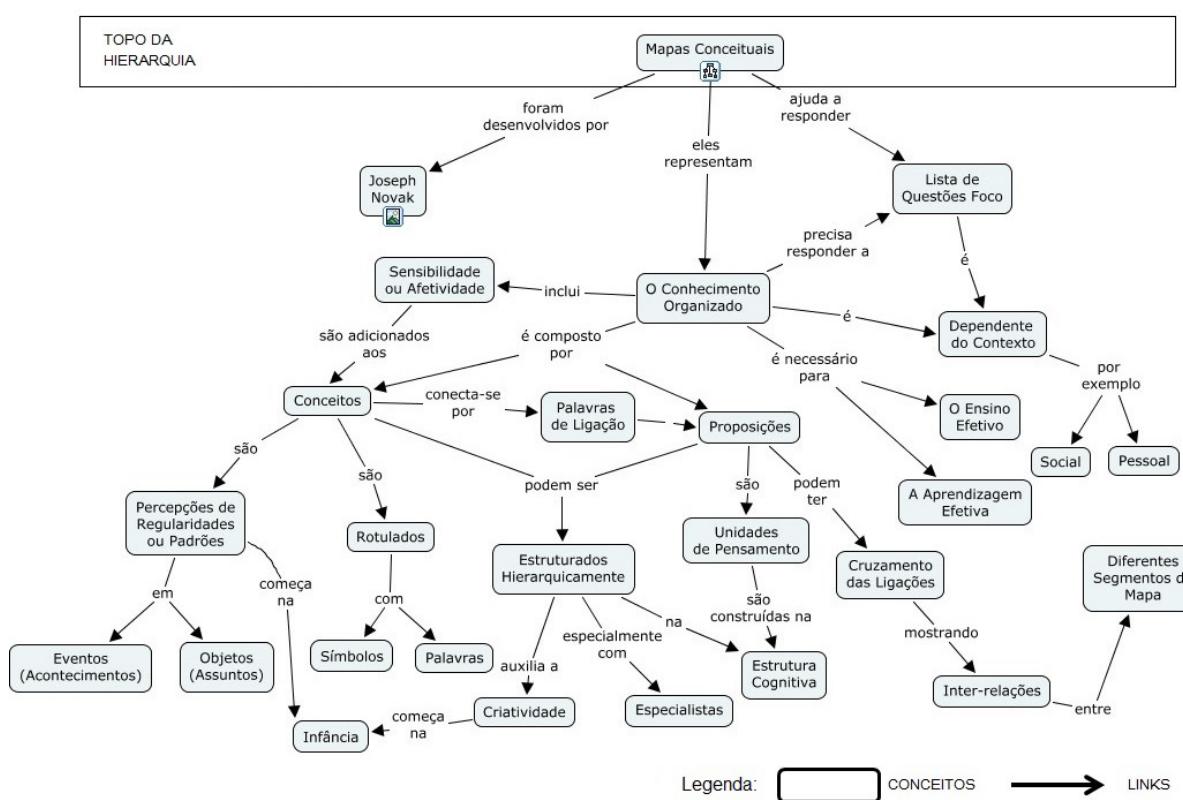


Figura 2: Exemplo de um MC
Fonte: Adaptado de (NOVAK; CAÑAS, 2008)

Considerando as características mencionadas, os MCs são representações bastante versáteis e por isso são utilizados em diversos contextos e áreas. Na Ciência da Computação, por exemplo, é possível identificar diferentes usos dos MCs. No estudo de Faily et al. (2012) os autores mostraram que os MCs poderiam ser utilizados para melhorar a qualidade da especificação

de requisitos de software. Também foram encontradas iniciativas do uso de MCs na modelagem de programas orientados a objeto (SIEN, 2010), extração e modelagem de requisitos (COFFEY et al., 2012; FAILY et al., 2012; KOF et al., 2007), segurança de softwares (SNIDER et al., 2014), melhorias nos processos de desenvolvimento de softwares (GURUPUR et al., 2014) e avaliação de softwares (WILLIAMS et al., 2012).

Visando otimizar o processo de seleção de estudos primários em um MS ou RS, Felizardo et al. (2017) propuseram um modelo de resumo gráfico utilizando MCs (Figura 3) que visa auxiliar os pesquisadores na criação de MCs que representem trabalhos científicos facilitando o entendimento dos leitores. O modelo proposto possui os conceitos dispostos de forma hierárquica, pois entende-se que esse formato gera uma melhor organização do conteúdo de estudos científicos (NOVAK; CAÑAS, 2008; ALIAS; SURADI, 2008). O modelo contém partes de um resumo estruturado representados pelos conceitos: título, contexto, objetivo, método, resultados e conclusão.

Os conceitos no modelo de Felizardo et al. (2017) são agrupados em níveis. O nível 0 contém o conceito principal (posicionado no centro do mapa), ao redor os demais conceitos relacionados ao tópico principal são desenhados e agrupados em níveis. O conceito no topo do mapa é o mais inclusivo. Em um nível mais baixo, mais detalhes são descritos, como ilustrado na Figura 3. O primeiro nível do modelo é composto de cinco conceitos fixos representando cada tópico dos resumos estruturados: Contexto, Objetivo, Método, Resultado e Conclusão. No segundo nível, cada conceito previamente definido pode ser refinado em novos conceitos: os conceitos opcionais. Por exemplo, o conceito “contexto” (nível 1) pode ser dividido em “área de pesquisa”, “problema”, e “motivação” (nível 2). Finalmente o terceiro nível contém conceitos variáveis que serão anexados pelo usuário do modelo.

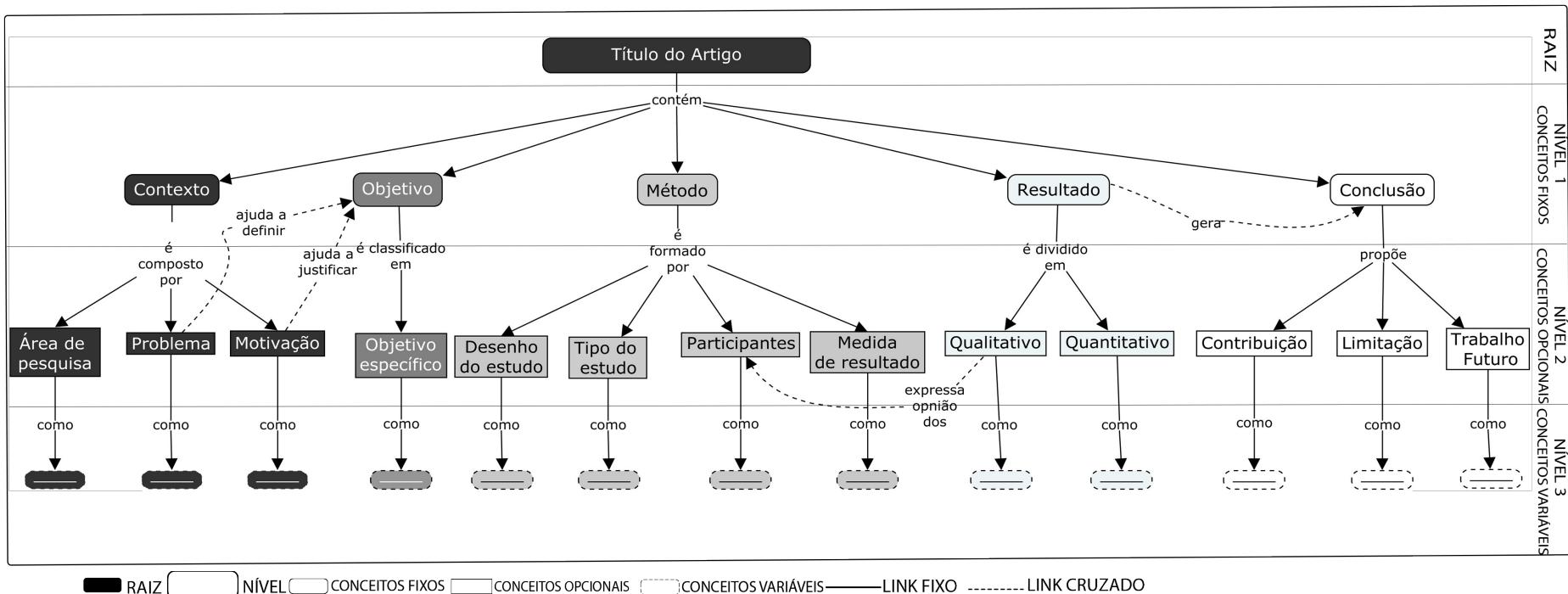


Figura 3: Estrutura proposta para representação de um resumo gráfico (TAKEMIYA; FELIZARDO, 2016)

Para melhor compreensão do modelo de Felizardo et al. (2017) os principais itens serão detalhados, a seguir:

Conceito principal: é a raiz do mapa, ou seja, é o nó no qual todos os outros nós irão se relacionar. O modelo definiu como nó principal título do estudo.

Conceitos fixos: correspondem as seções dos resumos estruturados e são obrigatórios no mapa.

Conceitos opcionais: são refinamentos dos conceitos fixos. No modelo apresentado (Figura 3) estão diretamente relacionados a condução dos estudos empíricos, tais como estudos experimentais. Os conceitos do nível 2 são opcionais, ou seja, cabe ao pesquisador analisar o tipo do estudo e optar pela utilização.

Conceitos Variáveis: são conceitos que devem ser preenchidos pelo pesquisador de acordo com o estudo.

Links básicos: definem o relacionamento entre os conceitos. Esses *links* devem relacionar conceitos que estão no mesmo ramo.

Links cruzados: ligam conceitos em diferentes ramos do MC. O uso dos links cruzados definidos pelo pesquisador reforçam a compreensão do conhecimento adquirido do estudo.

Takemiya e Felizardo (2016) buscam incentivar a utilização de resumos gráficos para auxiliar a etapa de seleção de estudos secundários na ES. No entanto, estes resumos são criados de forma manual. Como continuação do projeto desenvolvido por Takemiya e Felizardo (2016), este trabalho busca apoiar o processo de seleção de estudos primários a partir da construção automática dos MCs de forma automática aplicando técnicas de PLN.

2.3 PROCESSAMENTO DE LINGUAGEM NATURAL

O PLN estuda como projetar algoritmos que podem entender a semântica e sentimentos na linguagem escrita e falada, bem como responder de forma inteligente. O PLN é considerado um subconjunto da Inteligência Artificial (IA) considerando que a IA trata da construção de sistemas que podem tomar decisões inteligentes. O objetivo do PLN é fazer com que os computadores executem tarefas úteis envolvendo linguagem humana. Alguns benefícios da utilização de PLN é permitir a comunicação entre humanos e máquinas, melhorar a comunicação humano-humano ou simplesmente fazer um processamento útil de texto ou discurso (JURAFSKY; MARTIN, 2009).

O mecanismo da PLN envolve dois processos: Compreensão de Linguagem Natural

(CLN) e Geração de Linguagem Natural (GLN).

A CLN extrair informações de um determinado texto. Logo, para este processo a natureza e a estrutura de cada palavra de um texto devem ser entendidas. Para entender a estrutura do texto, a CLN tenta resolver vários tipos de ambiguidades presentes na linguagem natural: [Ambiguidade léxica] dado que as palavras têm múltiplos significados, [Ambiguidade sintática] uma frase pode ter múltiplas árvores de análise; [Ambiguidade semântica] frases podem ter múltiplos significados e [Ambiguidade anafórica] frases ou palavras mencionadas no texto podem ter um significado diferente de acordo com termos que foram mencionados anteriormente. Devido a ambiguidade, existem certas palavras diferentes com significado semelhante (sinônimos) e palavras com mais de um significado (polissemia).

O GLN é o processo de produção automática de texto a partir de dados estruturados em um formato legível com frases significativas. O GLN é dividido em três etapas propostas: planejamento de texto que é responsável pela ordenação do conteúdo em dados estruturados; planejamento de sentenças onde as frases são combinadas a partir de dados estruturados para representar o fluxo de informações; e finalmente a realização em que frases gramaticalmente corretas são produzidas para representar o texto.

Conforme Jurafsky e Martin (2009), existem 6 níveis de linguagem processadas pelas técnicas de PLN: (1) fonética e fonológica, relacionada à articulação de fonemas e uso em linguagem; (2) nível morfológico, relacionado a classes gramaticais (artigos, substantivos, verbos, entre outros) e formação de palavras (por raízes, prefixos, sufixos, entre outros); (3) nível sintático, busca encontrar dependências sintáticas entre frases em uma sentença, determinando funções sintáticas (sujeito, predicado, entre outros) e construindo árvores sintáticas; (4) nível semântico, possui um amplo conjunto de análises, como exemplo de desambiguação de palavras com mais de um significado (polissêmico); (5) pragmática, um nível preocupado com a informação implícita no texto; e (6) retórica, relacionado à estrutura do texto como um todo e as estratégias utilizadas para organização e criação de texto. Algumas técnicas de PLN se concentram em níveis específicos, como o reconhecimento de fala, enquanto outros abrangem vários níveis (como o resumo de documentos). Algumas tarefas são simples de resolver (exemplo: verificação ortográfica), enquanto outras podem ser extremamente desafiadoras (exemplo: tradução automática).

Para Cerqueira et al. (2010), as abordagens atuais da PLN se dividem basicamente em quatro categorias principais: simbólica, estatística, conexão e híbrida. A abordagem **simbólica** analisa os fenômenos linguísticos e seus paradigmas através de regras conhecidas da linguagem. Os métodos **estatísticos** utilizam cálculos matemáticos para gerar modelos e regras a

partir de exemplos de textos e sentenças. Os métodos de **conexionistas** também desenvolvem modelos, porém utilizam métodos estatísticos para complementar os métodos de representação de conhecimento. A última abordagem é denominada **híbrida**, que mescla métodos vindos de abordagens diferentes para suprir as necessidades do sistema ou mesmo reforçar os pontos fracos.

3 MAPAS CONCEITUAIS APLICADOS NA CIÊNCIA DA COMPUTAÇÃO

Esse capítulo apresenta uma investigação sobre aspectos associados ao uso de MCs na área da Ciência da Computação, que faz parte do primeiro objetivo desse projeto de mestrado. Para alcançar o objetivo proposto, um Mapeamento Sistemático (MS) foi conduzido e os principais aspectos investigados foram (SANTOS et al., 2017): (i) os propósitos da aplicação dos MCs na Ciência da Computação; (ii) as diferentes subáreas da Ciência da Computação na qual os MCs são aplicados; (iii) os recursos utilizados para construção dos MCs; (iv) as principais diretrizes utilizadas para avaliar os MCs; e (v) os benefícios e problemas reportados no uso dos MCs na Ciência da Computação. A seguir são apresentados brevemente o protocolo seguido no MS e os principais resultados.

3.1 PROTOCOLO DA REVISÃO

Para condução do MS o seguinte protocolo foi definido:

A. Questões de Pesquisa (QP)

Esse MS tem como objetivo responder as seguintes questões de pesquisa:

- (QP1) Quando e onde os estudos foram publicados?
- (QP2) Qual o propósito da aplicação dos MCs na Ciência da Computação?
- (QP3) Quais subáreas apresentaram o uso de MCs?
- (QP4) Quais são os principais instrumentos usados para desenvolver os MCs identificados?
- (QP5) Quais as diretrizes ou mecanismos usados para avaliar os MCs identificados?
- (QP6) Quais os benefícios relatados por autores relacionados com o uso de MCs?
- (QP7) Quais os problemas relatados por autores relacionados uso de MCs?

B. String de Busca

Na definição da *String* de busca foram consideradas duas áreas - “Mapa Conceitual” e “Ciência da Computação” (Tabela 1).

Tabela 1: String de busca - Aplicação dos MCs na Ciência da Computação

Área	Palavra-chave
<i>Concept Map</i>	“Concept Map”, “Concept Maps”, “Conceptual Map”, “Conceptual Maps”, “Concept Mapping”
<i>Computer Science</i>	“Computer Science”, “Computation”, “Software Engineering”, “Computer Engineering”, “Computing”
String de busca:	(“Concept Map” OR “Concept Maps” OR “Conceptual Map” OR “Conceptual Maps” OR “Concept Mapping”) AND (“Computer Science” OR “Computation” OR “Software Engineering” OR “Computer Engineering” OR “Computing”)

C. Bases de Busca

As bases de busca utilizadas no MS foram:

IEEE Xplore (<http://ieeexplore.ieee.org>)

ACM Digital Library (<http://dl.acm.org>)

ScienceDirect (<http://www.sciencedirect.com/>)

Scopus (<http://www.scopus.com/>)

Compedex (<http://www.engineeringvillage2.org>)

D. Critérios de Seleção

Os critérios de seleção foram organizados em um Critério de Inclusão (CI) e sete Critérios de Exclusão (CE).

(CI1) O estudo deve estar dentro da Ciência da Computação e discutir a iniciativa MCs.

(CE1) O estudo não possui um resumo;

(CE2) O estudo foi publicado somente como resumo;

(CE3) O estudo não está escrito em Inglês;

(CE4) O estudo é uma versão mais antiga de um estudo já considerado;

(CE5) O estudo não é um estudo primário, como editoriais, resumos de key-notes, workshops e tutoriais;

(CE6) O estudo não está no contexto da Ciência da Computação, embora apresenta o uso de MCs; e

(CE7) O estudo está no contexto da Ciência da Computação, mas não apresenta o uso de MCs.

3.2 SELEÇÃO DOS ESTUDOS

Após a execução da *string* de busca nas bases de buscas definidas, um total de 700 estudos foram retornadas. Foram considerados os estudos publicados até junho de 2016. O processo de seleção dos estudos foi dividido em cinco etapas, conforme apresentado na Figura 4.

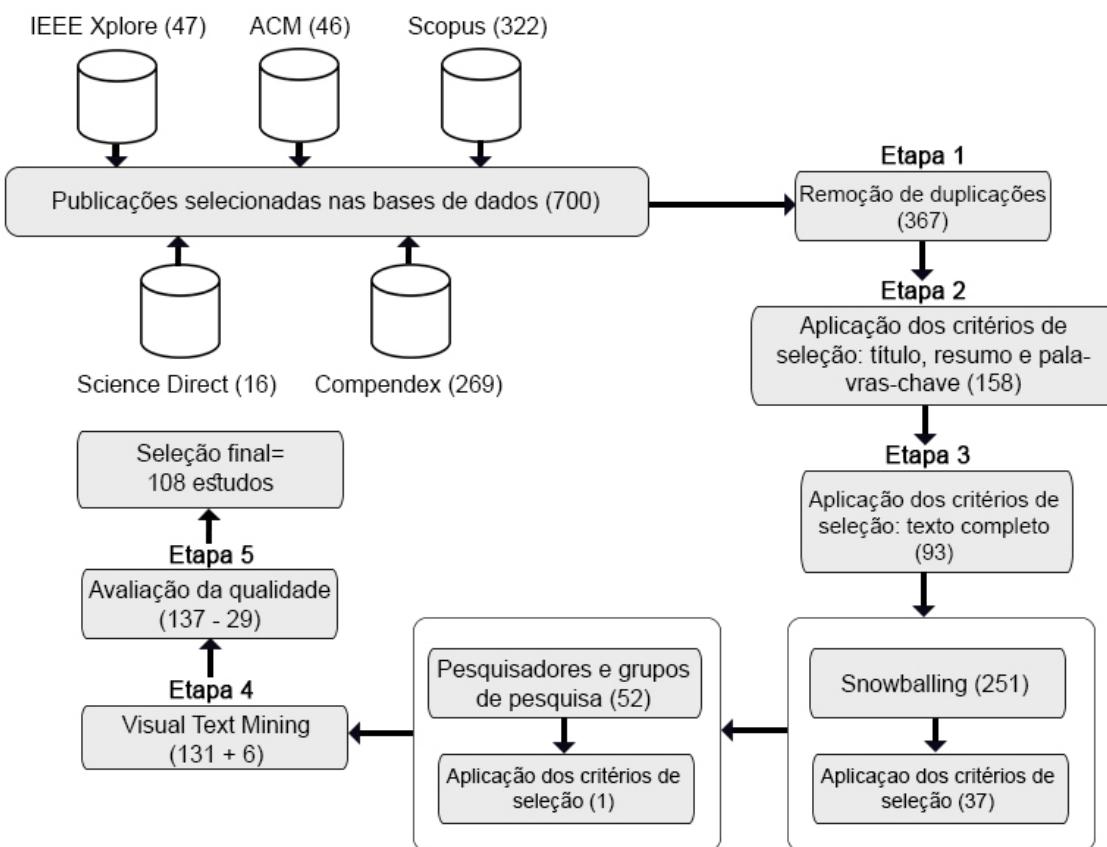


Figura 4: Fluxograma do MS - MCs na Ciência da Computação

Na primeira etapa do processo de seleção os estudos duplicados foram removidos. Foram eliminados aproximadamente 47,6% de estudos, restando 367 publicações para serem analisados na segunda etapa. Na segunda etapa, os critérios de seleção (critérios de inclusão e exclusão) foram aplicados no título, resumo e palavras-chave, resultando na seleção de 158 estudos (uma redução de aproximadamente 56,9%). Na terceira etapa, os critérios de seleção foram aplicados considerando o texto completo, resultando em um conjunto de 93 estudos (redução de 41,1%). Sobre estes 93 estudos considerados relevantes, foi aplicada a técnica de *snowballing*. *Snowballing* é um processo que verifica as citações de um determinado estudo, recupera os estudos, aplica os critérios de seleção e continua esse processo até não encontrar mais es-

tudos relevantes (FELIZARDO et al., 2016). Após a aplicação do *snowballing* 251 estudos foram identificados. A partir desses 251 estudos, os critérios de seleção foram aplicados considerando o título, resumo e palavras-chave. Em seguida, os critérios de seleção foram aplicados considerando o texto completo. Um total de 37 estudos resultaram desse estágio.

Ao fim do processo de *snowballing*, 130 estudos foram selecionados (93 das bases de dados e 37 do *snowballing*). A seguir, foi aplicada como técnica complementar a busca por grupos de pesquisa envolvidos na área. A busca foi realizada em páginas pessoais dos autores dos estudos selecionados até o momento. Um total de 52 estudos foram identificados. Após a aplicação dos critérios de seleção, apenas 1 estudo permaneceu. Como resultado, 131 estudos permaneceram (93 das fontes, 37 do *snowballing* e 1 da pesquisa direta a pesquisadores e grupos de pesquisa).

Na quarta etapa, foi aplicada a técnica de *Visual Text Mining* (VTM) para garantir que os estudos relevantes não fossem eliminados (FELIZARDO et al., 2012). A aplicação da técnica VTM teve como objetivo criar um mapa de conteúdo, isto é, uma representação visual bidimensional (2D) dos estudos selecionados (Figura 5). Cada estudo é representado graficamente como um elemento no plano, mostrado graficamente como um círculo. As posições dos estudos no mapa refletem as relações de similaridade entre seu conteúdo (título e resumo). Portanto, se os estudos são semelhantes (exibindo similaridade de conteúdo), eles são mostrados mais próximos uns dos outros; por outro lado, mais diferentes, mais distantes estão no gráfico. A partir da geração do mapa de conteúdo dos estudos do mapeamento, alguns estudos foram revisados a partir do seu posicionamento no mapa. Veja o exemplo do *Cluster 7* apresentado na Figura 5, esse estudo destacado foi revisado e sua classificação foi alterada de excluída para incluída. Um total de 6 estudos que foram inicialmente excluídos foram incluídos novamente após a revisão, totalizando 137 estudos.

Na quinta e última etapa da seleção, foi conduzida uma análise da qualidade dos estudos selecionados. Para analisar a qualidade dos estudos, foi considerada uma lista de verificação como um instrumento de qualidade, contendo sete questões (Tabela 2). A lista de verificação foi adaptada dos critérios genéricos de qualidade criados por Kitchenham (KITCHENHAM, 2004). Para cada pergunta na lista de verificação, a seguinte escala de pontuação foi aplicada: Sim - 1 ponto; Não - 0 pontos. O índice de qualidade total, portanto, variou no intervalo entre: 0-1 (muito pobre); 2 (justo); 3 - 4 (bom); 4 - 6 (muito bom) e 7 (excelente).

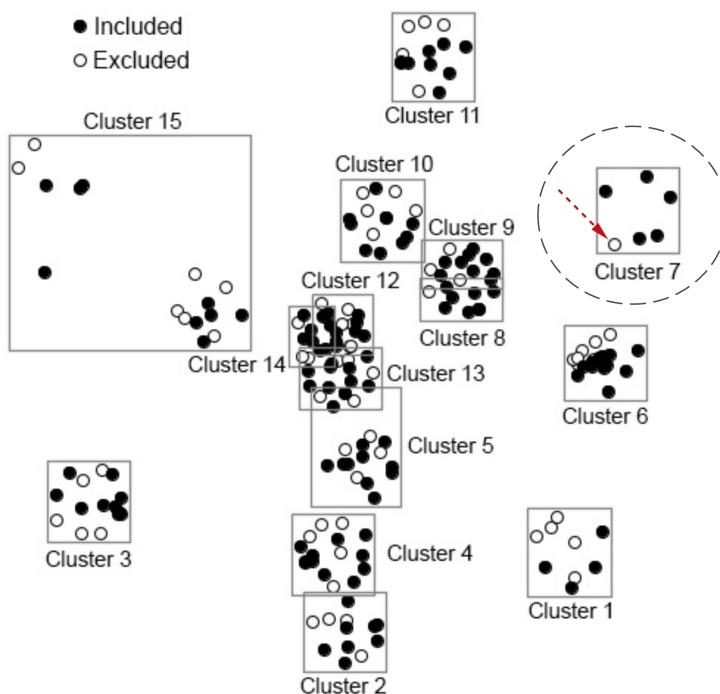


Figura 5: Mapa de conteúdo gerado a partir dos estudos selecionados.

Tabela 2: Checklist da avaliação de qualidade

Nº	Item	Resposta
Q1	O MC foi avaliado?	Sim/Não
Q2	Está claro como a forma de avaliação dos MCs foi aplicada?	Sim/Não
Q3	Está claro em qual tópico da Ciência da Computação os MCs foram aplicados?	Sim/Não
Q4	O estudo mostra melhorias na construção de MCs ou uma nova forma de construí-los?	Sim/Não
Q5	O MC foi construído com apoio de uma ferramenta?	Sim/Não
Q6	Estão claros os benefícios de utilizar MCs?	Sim/Não
Q7	Existem limitações no uso de MCs?	Sim/Não

O escore de qualidade foi utilizado para decidir a inclusão ou exclusão de um estudo. O ponto de corte para excluir estudos da lista de estudos selecionados foi 2 (muito pobre ou de qualidade justa). Um total de 29 estudos foram excluídos desta etapa. Como resultado final, a lista¹ contendo os 108 estudos a serem analisados estão disponíveis em (SANTOS et al., 2017).

3.3 RESULTADOS DO MAPEAMENTO

Nessa Seção serão apresentados brevemente os principais resultados identificados no MS. Detalhes sobre a condução e os resultados podem ser encontrados em (SANTOS et al.,

¹<https://goo.gl/9b6bZJ>

2017).

- **Frequência de publicação (QP1)**

A análise em relação à distribuição por ano dos 108 estudos (QP1) apresenta que a pesquisa sobre MCs na área de Ciência das Computação é recente. O interesse neste tópico cresceu moderadamente de 2006 até 2015, com um significante aumento do número de publicações em 2011 e 2012. Em relação aos veículos de publicação, as conferências podem ser consideradas o principal canal de comunicação utilizado, representando 54,62% dos estudos (59 estudos). As revistas ocuparam 35,1% (38 estudos), workshop 8,3% (9 estudos) e finalmente os simpósios com 1,8% (2 estudos).

- **Propósito do uso de MCs na Ciência da Computação (QP2)**

Pelo MS foi possível identificar o propósito da utilização dos MCs na Ciência da Computação. Os resultados mostram que os MCs são utilizados, principalmente, para auxiliar no ensino e aprendizado (63 estudos - 58,3%) e também na organização do conhecimento (40 estudos - 37,3%). Muitos estudos atribuídos nessas categorias usam MCs como ferramenta de ensino em cursos de graduação de Ciências da Computação e Engenharia de Software.

Os MCs também podem ser utilizados para determinar o nível de conhecimento de um aluno, ou seja, os MCs podem apresentar o quanto um aluno pode ter adquirido conhecimento. Um total de 19 estudos (aproximadamente 17,6%) têm como propósito a Avaliação do Nível de Aprendizagem. Os MCs são considerados ferramentas educacionais eficazes na sala de aula (FAILY et al., 2012; WILLIAMS et al., 2012; ATAPATTU et al., 2015a).

Fora do âmbito da educação, os MCs foram utilizados para a Representação Visual de Software (7 estudos - 6,5%). Por exemplo, Sien (2010) usou os MCs para representar o paradigma de orientação a objetos, em Sien (2010), Kof et al. (2007) e Coffey et al. (2012) os autores utilizaram MCs para representar requisitos de software, em Gurupur et al. (2014) os MCs foram utilizados para representar processos de desenvolvimento de software, e por fim, em Williams et al. (2012), os MCs foram utilizados na avaliação de softwares.

- **Subáreas aplicadas (QP3)**

Em relação à subárea da Ciência da Computação em que os MC têm sido aplicados, a maioria dos estudos (76 estudos - 70,3%) não deixam claro quais subáreas estão relacionados.

Os estudos discutem iniciativas com MCs de forma geral. Por exemplo, os estudos (ATAPATTU et al., 2015b; HUANG et al., 2012; Álvarez Bermejo et al., 2015) argumentam que o uso de MCs ajuda na conceitualização, esquematização e apresentação de informações relacionadas a computação em sala de aula, no entanto não definem exatamente quais informações.

Analizando a parcela de estudos que indicam qual a subárea na qual os MCs foram aplicados, duas categorias se destacam: Linguagens de Programação (10 estudos - 9,2%) e Engenharia de Software (12 estudos - 11,1%). Na categoria de linguagens de programação, os MCs foram usados principalmente para apresentar o ensino e aprendizado de paradigmas como a Programação Orientação a Objetos (POO) (BERGES et al., 2012; DOGAN; DIKBIYIK, 2016; HUBWIESER; MÜHLING, 2011; BERGES; HUBWIESER, 2013). Já em relação a Engenharia de Software, os MCs foram aplicados com o propósito de auxiliar as diferentes atividades que envolvem o processo de desenvolvimento de software, por exemplo, na especificação de requisitos ou na representação de rastreabilidade entre requisitos Faily et al. (2012), Williams et al. (2012), Coffey et al. (2012).

• Recursos utilizados para construção dos MCs (QP4)

A maioria dos estudos analisados utilizam abordagens próprias para construção dos MCs (58 estudos - 53,7%). Desses 58 estudos, 15 deles propõem ferramentas para a construção de MCs de forma colaborativa. As abordagens recentes de ensino e aprendizagem priorizam atividades que ocorrem de forma colaborativa. No estudo de Tomoto et al. (2011), por exemplo, é apresentada uma ferramenta colaborativa para integração e compartilhamento de conhecimento usando a construção de MCs colaborativamente. O estudo de Blebic et al. (2007) apresenta a ferramenta *MaGIA*, um sistema multiusuário projetado para a construção colaborativa de MCs. Martinez-Maldonado et al. (2015) apresenta o *Tabletop Supported Collaborative Learning* (TSCL), um modelo que proporciona suporte para construção MCs a partir de sistemas *Tabletop* e que os usuários interagem a partir de um processo colaborativo.

Outra ferramenta citada pelos estudos selecionados é a *Concept Map Editor CmapTools*². A *CmapTools* é uma ferramenta bastante popular e foi utilizada em 34 estudos (31,4%) na construção de MCs.

Mesmo com menor representatividade, outros recursos são mencionados para construção de MCs, por exemplo, a criação de MCs baseado em uma ontologia existente; extração de conceitos e relacionamentos a partir de texto para gerar MCs automaticamente, denominado Mineração de Mapas Conceituais - MMC; a utilização de *parser* baseado em *Extensible Markup*

²<http://cmap.ihmc.us/>

Language (XML) para extrair conceitos e as suas relações; e o tradicional método de construir MCs usando papel e caneta.

- **Avaliação dos MCs (QP5)**

Uma das principais questões relacionadas aos MCs é como avaliá-los. Os estudos mostraram que existem vários métodos para avaliação, um exemplo de método é a atribuição de pontos baseados na avaliação humana. A análise apresentada em Graudina e Grundspenkis (2011) mostra que a maioria dos métodos são baseadas em medidas quantitativas, ou seja, análise estrutural (número de proposições válidas, níveis de hierarquia, etc). Outro método de avaliação é a comparação dos MCs criados com um ou mais mapas de especialistas da área. Alguns autores também propuseram formas automáticas de avaliação de MCs baseada em um software de comparação.

Foi constatado que dos 108 estudos deste mapeamento, apenas 38 (35,1%) apresentaram algum tipo de avaliação. Depois de classificar os principais mecanismos usados para avaliar MCs identificados no estudo, três tipos de avaliação foram identificados. (I) Comparar com mapa de especialista; (II) Análise estrutural; (III) Suporte de software para comparação.

A primeira classificação está relacionada a comparação de MCs com um mapa elaborado por um especialista, o qual é usado como referência para corrigir outro criado posteriormente. Dos 38 estudos que avaliaram os MCs criados, 23 (84,2%) utilizaram essa classificação. Em Rigby et al. (2009), por exemplo, foram examinadas as formas de otimizar o aprendizado em um curso de Ciência da Computação. Os estudantes criaram MCs do seu entendimento de assuntos abordados no curso, os MCs foram avaliados de acordo com o MC criado por um professor. Em Qasim et al. (2013) foi proposto um algoritmo para a extração automática de MCs a partir de textos. Os mapas criados foram verificados por especialistas com doutorado em Sistemas de Informação.

A segunda classificação refere-se a análise estrutural dos MCs. Essa classificação considera o número de proposições válidas, níveis de hierarquia, número de conceitos, etc. Em Gurupur et al. (2015), foi apresentada uma ferramenta que pode avaliar o aprendizado dos alunos utilizando MCs e análises de Cadeias de Markov. Os MCs criados pelos alunos são comparados com outros a partir do cálculo do número de conceitos na hierarquia. Em Arruarte et al. (2014), um estudo foi conduzido utilizando a construção de MCs de forma colaborativa para melhorar o aprendizado de alunos do curso de Engenharia da Computação. Neste estudo os MCs foram avaliados a partir da média de número de conceitos, relacionamentos e integralidade dos mesmos.

A terceira classificação é a avaliação de MCs com suporte de software. Em Zvacek et al. (2012), por exemplo, é apresentado um processo para analisar MCs criados por estudantes sobre a estrutura de um computador. O software implementado mescla todos os MCs criados para extrair quais relações foram perdidas. Outro exemplo é apresentado em Vaishnavi et al. (2016), em que os MCs desenvolvidos por estudantes são convertidos em formato XML e os conceitos e relacionamentos são identificados. A avaliação do nível de entendimento sobre um assunto é analisado pela curva de conhecimento gerada por esta abordagem.

- **Benefícios e Problemas reportados no uso dos MCs (QP6 e QP7)**

Foram destacados nos estudos incluídos diversos benefícios em utilizar MCs na Ciência da Computação:

- **Estrutura flexível e fácil de entender:** os MCs são uma forma simples de representação e fácil de entender a conexão entre dois itens.
- **Representação de conhecimento:** O conhecimento representado no formato visual torna mais fácil o reconhecimento e a compreensão das informações recebidas. Em muitas disciplinas, várias formas de MCs já são usadas como sistemas formais de representação de conhecimento, por exemplo: *semantic networks* em Inteligência Artificial, *bond graphs* em engenharia mecânica e elétrica, redes de Petri em comunicações, requisitos no processo de desenvolvimento de software.
- **Compartilhamento de conhecimento:** pesquisadores investigaram o compartilhamento de conhecimento por meio dos MCs. A integração semântica recebeu uma atenção significativa nos últimos anos, particularmente de pesquisadores de bases de dados que procuram facilitar o compartilhamento de informações e integrar fontes de dados heterogêneas (Kristensen et al., 2009).
- **Contribuição Pedagógica:** na área educacional os MCs podem ser usados para auxiliar o aprendizado e também para avaliar o conhecimento (ARRUARTE et al., 2008). É possível notar a partir dos resultados do mapeamento que existe um conjunto de estudos que avaliam o efeito dos MCs nos estudantes em uma ampla diversidade de áreas, especialmente na Ciência da Computação. De acordo com Gurupur et al. (2015), os MCs fornecem informações visuais para os instrutores do nível de conhecimento dos estudantes. Assim, os MCs podem auxiliar o desenvolvimento de certas habilidades no aluno, por exemplo, identificar os conceitos chave de um domínio, sintetizar ideias transformando-as em informações e memorizar conceitos de uma área de estudo.

Apesar do uso de MC trazer muitos benefícios, alguns autores, embora poucos, também destacam os problemas que podem surgir, tais como:

- **Dificuldade em criar os MCs:** visto que o conhecimento do domínio que será trabalhado é muito importante para assegurar a qualidade do MC.
- **Sobrecarga cognitiva:** considerando a memória de curto prazo, os MCs podem causar sobrecarga cognitiva quando eles contém muitos conceitos em sua estrutura. Quando um MC possui muitos conceitos e relacionamentos, pode se tornar difícil de entender o conteúdo do mapa (WILLIAMS et al., 2012).
- **Aumento da carga de trabalho:** É importante destacar que a construção de MCs demanda tempo, assim, existe um risco potencial em utilizar MCs quando o usuário precisa cumprir prazos. Por exemplo, no desenvolvimento de software em empresas, a construção de MCs juntamente com outros artefatos podem gerar sobrecarga de trabalho. Alguns estudos ainda mostram que a automatização da construção dos MCs a partir de texto é uma tendência de pesquisa investigada atualmente (QASIM et al., 2013; CHEN et al., 2008).
- **Escalabilidade:** os MCs podem se tornar grandes rapidamente e muito complexos para serem entendidos pelo usuário. A representação de um MC pode perder sua utilidade quando a quantidade conceitos e links entre os conceitos crescem.

3.4 CONSIDERAÇÕES FINAIS

A principal contribuição do MS apresentado foi evidenciar como os MCs foram utilizados na Ciência da Computação. Este MS mostrou que os MCs foram amplamente utilizados com diferentes finalidades e em diferentes subáreas, por exemplo, a Engenharia de Software. No entanto, o período em que os estudos foram publicados demonstra que o interesse em MC na ES pode ser considerado recente. Por exemplo, dos 11 artigos enquadrados na categoria de ES, 8 foram publicados a partir de 2009 e três em 2013.

Um resultado que chamou a atenção foi em relação à avaliação dos MCs. Embora existam na literatura diversos métodos de avaliação de MCs, apenas 35,1% dos MC incluídos neste mapeamento foram avaliados. A minoria dos estudos incluídos neste mapeamento (38 de 108) relatou alguma avaliação de MC.

Muitos pesquisadores que realizam estudos secundários, como uma de suas primeiras atividades, precisam ler resumos de estudos para decidir sobre sua inclusão ou exclusão. Em alguns domínios, como a medicina clínica, a educação, a psicologia e as ciências sociais,

a revisão dos resumos é mais que suficiente para determinar sua relevância (KITCHENHAM, 2004). No entanto, em Ciência da Computação, especialmente, na Engenharia de Software, há resumos não estruturados nos estudos. Isso torna difícil identificar a relevância ou não dos estudos (BRERETON et al., 2007). A literatura recente forneceu evidências de que resumos não estruturados e mal escritos podem comprometer a atividade de seleção (BRERETON et al., 2007; DYBA et al., 2007; KITCHENHAM et al., 2008; DYBA; DINGSOYR, 2008; PETERSEN; NAUMAN, 2011; ZHANG; MUHAMMAD, 2011; ZHANG; MUHAMMAD A. B. TELL, 2011; ZHANG; MUHAMMAD, 2012).

Uma solução potencial para minimizar o problema da morosidade no processo de seleção de estudos primários é promover o uso de resumos estruturados e gráficos (KITCHENHAM et al., 2008). Como os MCs podem ajudar na compreensão de um estudo, um dos objetivos desse trabalho é investigar a criação automática de resumos gráficos para apoiar a atividade de seleção em estudos secundários usando MCs (FELIZARDO et al., 2016).

4 CONSTRUÇÃO DE MAPAS CONCEITUAIS A PARTIR DE PLN

É apresentado nesse capítulo o segundo Mapeamento Sistemático (MS) conduzido, o qual também faz parte do primeiro objetivo desse projeto de mestrado. Nesse MS buscou-se sumarizar as principais iniciativas de construção de MCs a partir de técnicas de PLN.

4.1 PROTOCOLO DE REVISÃO

Para condução do MS sobre construção de MC a partir de PLN, o seguinte protocolo foi definido:

A. Questões de Pesquisa (QP)

Esse MS tem como objetivo responder as seguintes questões de pesquisa:

- (QP1) Qual o propósito da construção dos MCs?
- (QP2) Qual o nível de automatização que foi utilizado para construção dos MCs?
- (QP3) Quais abordagens foram utilizadas para construção dos MCs?
- (QP4) Quais ferramentas foram utilizadas para construção dos MCs?
- (QP5) Quais algoritmos foram utilizados para construção dos MCs?

B. String de Busca

Na definição da *string* de busca foram consideradas duas áreas - “Mapa Conceitual” e “Processamento de Linguagem Natural” (Tabela 1).

C. Bases de Busca

As bases de busca utilizadas no MS foram:

- IEEE Xplore (<http://ieeexplore.ieee.org>)
- ACM Digital Library (<http://dl.acm.org>)
- ScienceDirect (<http://www.sciencedirect.com/>)
- Scopus (<http://www.scopus.com/>)

Tabela 3: String de busca - Construção de MCs utilizando PLN

Áreas	Palavras-chave
<i>Concept Map</i>	“concept map”, “concept maps”, “conceptual map”, “concept map mining”, “conceptual maps”, “concept mapping”.
<i>Natural Language Processing</i>	“text data mining”, “text mining”, “text analysis”, “automatic”, “automatically”, “text processing”, “natural language processing”, “NLP”.
<i>Search String</i>	(“concept map” OR “concept maps” OR “conceptual map” OR “concept map mining” OR “conceptual maps” OR “concept mapping”) AND (“text data mining” OR “text mining” OR “text analysis” OR “automatic” OR “automatically” OR “text processing” OR “natural language processing” OR “NLP”)

D. Critérios de Seleção

Os critérios de seleção foram organizados em um Critério de Inclusão (CI) e sete Critérios de Exclusão (CE).

- (CI1)** O estudo deve apresentar uma abordagem da construção MCs a partir de PLN.
- (CE1)** O estudo não possui um resumo;
- (CE2)** O estudo foi publicado somente como resumo;
- (CE3)** O estudo não está escrito em Inglês;
- (CE4)** O estudo é uma versão mais antiga de um estudo já considerado;
- (CE5)** O estudo não é um estudo primário, como editoriais, resumos de key-notes, workshops e tutoriais; e
- (CE6)** O estudo apresenta uma abordagem de construção de MCs, mas não usa PLN.

4.2 SELEÇÃO DE ESTUDOS

Conforme apresentado na Figura 5, o MS foi dividido em duas etapas; revisão manual (A) e revisão automática (B). O objetivo da revisão manual foi identificar as principais palavras-chave para construir a *string* de busca do mapeamento. Na revisão manual foram utilizados todos os volumes dos Anais da Conferência Internacional de Mapas Conceituais (CMC). (CMC¹) até 2016 (Tabela 4). Os anais da CMC contém iniciativas em múltiplas áreas focando em novas maneiras de usar e construir MCs.

Ao analisar os estudos publicados na CMC, 52 estudos foram considerados estudos candidatos (Figura 5 - Parte A). Na primeira etapa, os critérios de seleção foram aplicados

¹<http://cmc.ihmc.us/cmc-proceedings/>

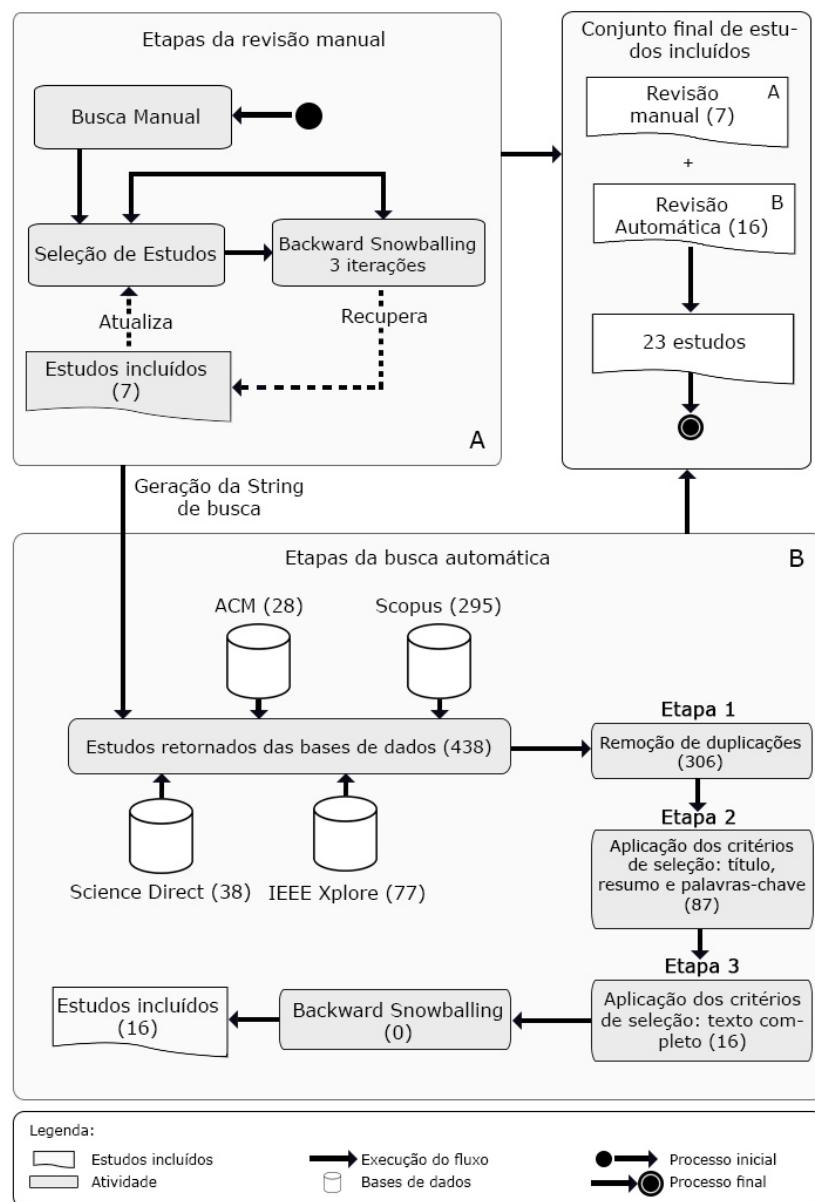


Figura 5: Processo de seleção conduzido no MS

considerando o título, resumo e palavras-chave. Ao final dessa etapa 20 estudos foram selecionados. A seguir, foram aplicados os critérios de seleção considerando o texto completo, levando a incluir 6 estudos. Sobre esses 6 estudos, foi utilizada a técnica de *backward snowballing*. O processo foi realizado com três iterações. Depois de aplicar os critérios de seleção, apenas 1 estudo permaneceu. Assim, o conjunto final foi composto por 7 estudos, sendo 6 estudos retirados dos anais da Conferência do CMC e 1 estudo encontrado por meio do processo de *snowballing*. A partir do conjunto de estudos coletado na revisão manual, foram identificadas as principais palavras-chave para compor o conjunto de termos usados na revisão automática. Os termos que compõem a *String de busca* são apresentados na Tabela 3.

Tabela 4: Anais da Conferência Internacional de MCs usado na busca manual

Ano	Volume	Tema
2004	1	Mapas Conceituais: Teoria, Metodologia, Tecnologia
2006	2	Mapas Conceituais: Teoria, Metodologia, Tecnologia
2008	3	Mapas Conceituais: Conectando Educadores
2010	4	Mapas Conceituais: Tornando o aprendizado representativo
2012	5	Mapas Conceituais: Teoria, Metodologia, Tecnologia
2014	6	Uso de Mapas Conceituais para aprender e inovar
2016	7	Inovando com Mapas Conceituais

No processo de revisão automática, foram considerados os estudos publicados até dezembro de 2016. Como resultado, um total de 438 publicações foram coletadas, sendo: 77 da base IEEE *Xplore*, 295 da *Scopus*, 38 da *Science Direct*, 28 da Biblioteca Digital ACM (Figura 5 - Parte B). O processo de seleção na revisão automática foi dividido em 3 etapas. Na primeira etapa, os estudos duplicados foram eliminados reduzindo o número de publicações para 306 (aproximadamente 29,5% de redução). Na segunda etapa, os critérios de seleção foram aplicados considerando o título, resumo e palavras-chave. Foram eliminadas 219 publicações (71,1%). Embora algumas publicações citem no resumo os termos pesquisados, muitos deles não apresentam uma abordagem de construção MCs usando PLN. Na terceira etapa, os critérios de seleção foram aplicados considerando o texto inteiro, resultando em uma redução de 61,7%. A partir da terceira etapa do processo de seleção, 16 estudos foram considerados relevantes. 16 estudos foram selecionados de um conjunto inicial de 438, apresentando então uma taxa de redução de cerca de 96,3%. Com estes resultados, foram selecionados 23 estudos a serem analisados (7 da revisão manual e 16 de revisão automática). Os 23 estudos considerados no conjunto final do mapeamento estão apresentados na Tabela 5.

Tabela 5: Lista de IDs e referências

ID	Citation	ID	Citation
1	(AGUIAR; CURY, 2016)	13	(OLNEY et al., 2011)
2	(BUI et al., 2016)	14	(QASIM et al., 2013)
3	(CAPUTO; EBECKEN, 2011)	15	(RICHARDSON; FOX, 2005)
4	(CEAUSU; DESPRÉS, 2007)	16	(VALERIO et al., 2012)
5	(CHAMPAIGN; MCCALLA, 2015)	17	(VILLALON; CALVO, 2011)
6	(SILVA et al., 2014)	18	(WANG et al., 2008b)
7	(JUNQUEIRA et al., 2014)	19	(WANG et al., 2009)
8	(KE, 2013)	20	(YANG, 2015)
9	(KOWATA et al., 2010)	21	(YOON et al., 2014)
10	(MARTIN-RODILLA; GONZALEZ-PEREZ, 2016)	22	(ZOUAQ; NKAMBOU, 2008)
11	(NUGUMANOVA et al., 2015)	23	(ZOUAQ et al., 2007)
12	(OLIVEIRA et al., 2001)		

4.2.1 CLASSIFICAÇÃO DOS ESTUDOS SELECIONADOS

Para responder as questões de pesquisa com base nos 23 estudos selecionados, foi utilizado um formulário contendo os seguintes parâmetros: referência bibliográfica, propósito, nível de automação, técnicas, ferramentas e algoritmos usados para construção de MCs. Antes da extração dos resultados, as categorias para classificar os estudos foram definidas de acordo com as questões de pesquisa. As classificações definidas foram:

Classificação para o propósito dos estudos (QP2): nesta classificação, buscou-se entender os propósitos de construir MCs automaticamente. Foram identificadas duas características principais que contém os seguintes propósitos: educacional e analíticos. Educacional: os MCs são utilizados como ferramenta para melhorar o ensino e o processo de aprendizado. Analítico: os MCs são utilizados para explorar, analisar e descobrir padrões em bases de dados específicas.

Classificações para o nível de automação (QP3): Duas categorias foram utilizadas para classificar o nível de automação para construir MCs: automática e semiautomática. Automática: é uma iniciativa onde a entrada é feita (em linguagem natural) e o MC é construído automaticamente sem interferência humana. Semiautomática: é bastante similar a categoria automática, no entanto ela pode sofrer interferência humana na construção dos MCs.

Classificação para as técnicas usadas (QP4): as técnicas de PLN podem ser classificadas em três principais categorias: linguística, estatística e híbrida. Linguística: usa a morfologia, gramática, sintaxe, semântica e outros recursos linguísticos para lidar com texto. Os métodos usados são diretamente ligados a estrutura baseada na linguística e reconhecimento de padrão. Estatístico: nesta categoria é bastante comum analisar a frequência dos termos e co-ocorrência dos termos no documento. Uma das vantagens de usar esta categoria é a simplicidade por não considerar a linguística. No entanto, existem também desvantagens, assim como a perda de conteúdo relevante de informação quando comparado com a iniciativa linguística. Híbridos: são estudos que usam a combinação de técnicas linguísticas e estatísticas, baseadas no *parsing* sintático, filtros linguísticos e medidas estatísticas.

Classificação para ferramentas utilizadas (QP5): essa classificação discute as ferramentas que foram usadas ou propostas para auxiliar o processo de construção dos MCs a partir de linguagem natural. As ferramentas encontradas podem agir em qualquer etapa da construção dos mapas, ou seja, na extração de conceitos, relacionamentos ou até mesmo a sumarização e exibição dos dados.

Classificação para os algoritmos usados (QP6): nessa classificação foram conside-

dados todos os algoritmos de PLN usados para realizar a extração de MCs.

4.3 RESULTADOS DO MAPEAMENTO

A seguir serão apresentados os resultados obtidos com a análise dos estudos coletados pelo MS. A Figura 6 apresenta uma visão geral dos esforços na área de construção de MCs usando PLN. Assim como a Figura 6 sugere, o uso de PLN para construção de MCs é recente, ocorrendo basicamente de 2001 até os dias atuais, apresentando um crescimento no número de publicações em 2011, 2014 até 2016.

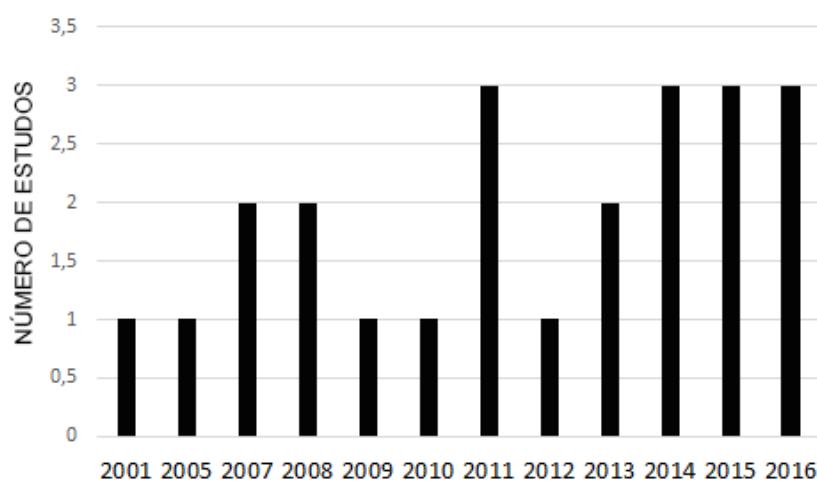


Figura 6: Distribuição dos estudos selecionados ao longo dos anos

Os estudos selecionados foram publicados em quatro principais veículos: *Journals*, conferências, *Workshops* e simpósios. As conferências têm sido o principal veículo de apresentação dos trabalhos de MCs e PLN, 56,5% (13 estudos de 23). Os *Journals* publicaram 34,7% dos estudos selecionaram (8 de 23). Finalmente, os *Workshops* e Simpósios publicaram 1 estudo cada (1 de 23 estudos), representando então 4,3% dos estudos.

- **Propósito do estudo**

O propósito da criação de MCs nos estudos selecionados é analisado na QP2. A Figura 7 apresenta que os fins educacionais têm a maior representatividade nas iniciativas analisadas. Estudos classificados como educacionais (14 estudos - 60,8%) mostraram interesse em usar o conhecimento extraído dos textos para melhorar o ensino de uma determinada área. Por exemplo, em Yang (2015), os efeitos da construção automática e da avaliação de MCs com três camadas foram avaliados. Os autores buscaram investigar se a utilização dos MCs poderia melhorar a escrita de resumos por estudantes universitários. Por outro lado, estudos classificados

como analíticos (9 estudos - 39,3%) usaram MCs como ferramentas para entender uma domínio específico. Em Caputo e Ebecken (2011), por exemplo, os autores interpretam a informação presente nos documentos que descrevem os produtos e usam os MCs para criar um mecanismo para acelerar a interpretação das informações.

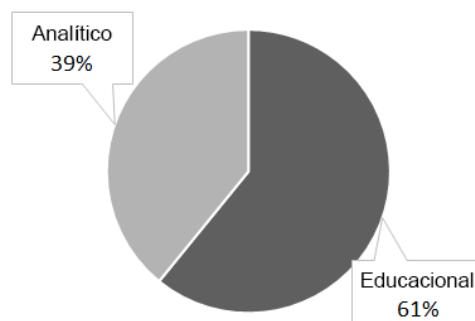


Figura 7: Propósito de aplicação dos estudos

- **Nível de automatização na criação de MCs**

A QP3 busca analisar o nível de automatização ao criar MCs. A automatização dos MCs foi classificada em dois possíveis níveis: construção automática ou semiautomática. A Figura 8 mostra que a maioria dos estudos usam a extração automática dos termos e relacionamentos (17 estudos - 73,9%). Em Qasim et al. (2013), por exemplo, os autores propuseram um algoritmo para extração automática de termos candidatos a partir de texto. Técnicas como resolução de anáforas (para pronomes para aumentar o *recall* ao encontrar mais proposições em documentos de texto), *clustering*, extração de informação e mensuramento a partir da similaridade estrutural foram aplicadas para atingir um alto desempenho.

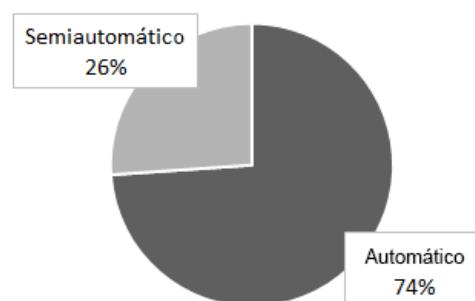


Figura 8: Nível de automatização dos estudos

Por outro lado, estudos classificados como semiautomáticos representam 26,8% (6 estudos). Esse nível de automatização possuí algumas intervenções humanas na construção de

MCs. Em Oliveira et al. (2001), os autores propuseram a construção de MCs a partir de PLN e aprendizado de máquinas. Entretanto, em algum ponto da construção, o sistema desenvolvido interage com os usuários realizando questões para validar os conceitos ou relacionamentos e este procedimento é capaz de melhorar o desempenho do software.

- **Técnicas utilizadas para construção de MCs**

Na QP4 foram mapeados os tipos de técnicas de PLN usadas nos estudos selecionados. As técnicas utilizadas para construção de MCs foram classificadas em linguísticas, estatísticas e híbridas. A Figura 9 apresenta a distribuição dos estudos de acordo as técnicas. O uso de técnicas estatísticas foi identificado em 17,3% (4 estudos) dos estudos selecionados. Esse estudo foram baseados em algoritmos clássicos, como a frequência de termos (*Term Frequency*) (CAPUTO; EBECKEN, 2011; NUGUMANOVA et al., 2015; YANG, 2015). Outras técnicas amplamente utilizadas são as linguísticas (9 estudos - 39,1%). Elas têm como principal característica o uso de regras gramaticais e linguísticas para extração de conceitos e relacionamentos. Alguns estudos usam a classificação de palavras de acordo com a sua função em uma sentença (*Part-of-Speech Tagging*) (WANG et al., 2009; YANG, 2015; ZOUAQ; NKAMBOU, 2008). Outro exemplo de técnica linguística aplicada nos estudos é a extração dos relacionamentos mapeando o sujeito, verbo e objetos para determinar a existência da relação (CHAMPAGN; MCCALLA, 2015). Finalmente, as técnicas chamadas Híbridas (10 Estudos - 43,5%) combinam algoritmos estatísticos e linguísticos na construção de MCs. A coleta de conceitos através da frequência de termos pode ser combinada com técnicas que localizam a função de cada palavra em uma sentença, resultando em uma abordagem mais rica, como apresentado em (WANG et al., 2008a).

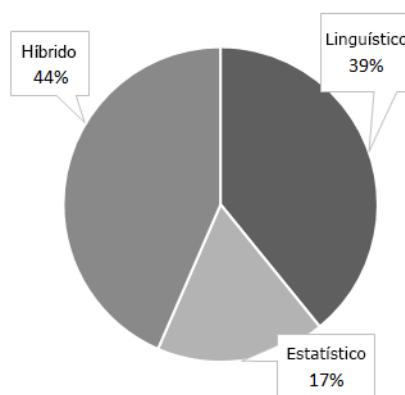


Figura 9: Técnicas utilizadas para construção de MCs

Para sumarizar o resultado encontrado na QP2, QP3, QP4, os dados foram adicionados em formato de Tabela (Tabela 6). Desta forma, é possível visualizar exatamente quais estudos

são classificados em cada categoria criada para cada QP.

Tabela 6: Classificações e estudos encontrados pelas QP2, QP3 e QP5

ID	Eduacional	Analítico	Automático	Semiautomático	Linguístico	Estatístico	Híbrido
(AGUIAR; CURY, 2016)		✓	✓		✓		
(BUI et al., 2016)	✓			✓			✓
(CAPUTO; EBECKEN, 2011)		✓	✓				✓
(CEAUSU; DESPRÉS, 2007)		✓	✓				✓
(CHAMPAIGN; MCCALLA, 2015)	✓		✓				✓
(SILVA et al., 2014)		✓	✓		✓		
(JUNQUEIRA et al., 2014)	✓		✓				✓
(KE, 2013)	✓		✓		✓		
(KOWATA et al., 2010)		✓	✓		✓		
(MARTIN-RODILLA; GONZALEZ-PEREZ, 2016)		✓		✓	✓		
(NUGUMANOVA et al., 2015)	✓		✓			✓	
(OLIVEIRA et al., 2001)	✓			✓			✓
(OLNEY et al., 2011)	✓		✓		✓		
(QASIM et al., 2013)		✓	✓				✓
(RICHARDSON; FOX, 2005)	✓			✓	✓		
(VALERIO et al., 2012)	✓		✓		✓		
(VILLALON; CALVO, 2011)	✓		✓				✓
(WANG et al., 2008b)		✓	✓		✓		
(WANG et al., 2009)		✓	✓		✓		
(YANG, 2015)	✓		✓			✓	
(YOON et al., 2014)		✓	✓				✓
(ZOUAQ; NKAMBOU, 2008)	✓		✓		✓		
(ZOUAQ et al., 2007)	✓			✓		✓	

• Ferramentas para construção dos MCs

As ferramentas (QP5) usadas para construção de MCs são apresentadas na Figura 10. Neste contexto foram encontrados 15 ferramentas diferentes. A *Wordnet* foi utilizada em 6 estudos (26,8%), a maioria deles aplicando para normalizar os sinônimos para que seja possível lidar com um grande número de termos que variam de acordo com o contexto (WANG et al., 2008a). Outra ferramenta amplamente usada foi o *Stanford NLP Parser* (6 estudos - 26,8%). Esta ferramenta têm módulos que implementam técnicas de PLN, incluindo a identificação de partes da fala (part-of-speech), (ZOUAQ; NKAMBOU, 2008; ZOUAQ et al., 2007). Similarmente, o *Natural Language Toolkit* (3 estudos - 13,4%) foi também utilizado para extrair conceitos e relacionamentos usando algoritmos estatísticos e linguísticos (CAPUTO; EBECKEN, 2011). Outras ferramentas foram identificadas em 12 artigos: KEA, Tex Comom, WEKA, entre outras.

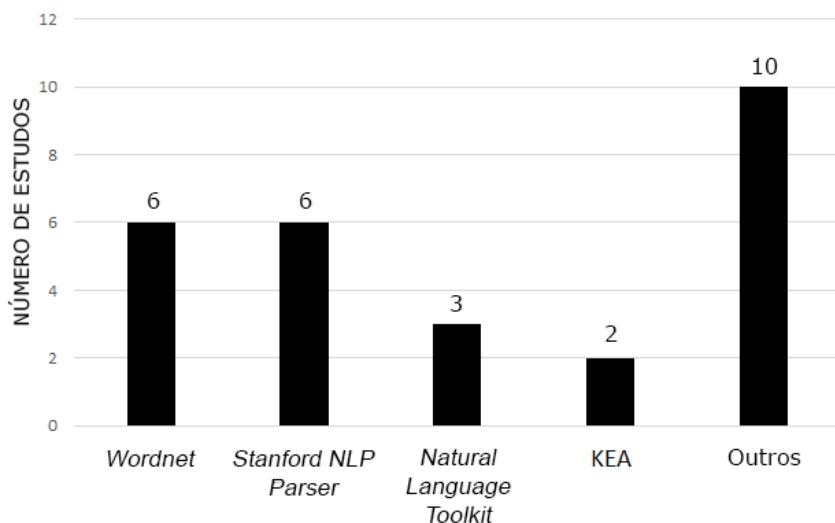


Figura 10: Ferramentas para construção de MCs

- **Algoritmos para construção de MCs**

Para responder a QP6, foram mapeados os algoritmos usados para auxiliar a criação de MCs e extração de conceitos e relacionamentos em um texto. Assim como apresentado na Figura 11, existe um grande número de algoritmos usados para gerar MCs. O algoritmo mais popular é o *Part-Of-Speech Tagger* (POS tagger) (14 estudos - 60,8%), responsável no reconhecimento das partes da fala. O segundo algoritmo mais utilizado foi o *Term Frequency* que atua contando a frequência de termos. Muitas vezes este algoritmo é utilizado em conjunto com uma variação que considera o inverso da frequência do termo em um conjunto de documentos (*TF-Inverse Document Frequency*) (7 estudos - 30,4 %). Foram encontrados também alguns algoritmos linguísticos, assim como o Sujeito-Verbo-Objeto (4 estudos- 17,3%) e resolução de anáforas (5 estudos 21,7%), estes também são aplicados para extrair conceitos e relacionamentos. Outros algoritmos são mencionados em outros 9 estudos: *Latent Semantic Analysis*, *Clustering*, entre outros.

Os estudos classificados em cada categoria associada em QP5 e QP6 foram sumarizados na Tabela 7.

Tabela 7: Relação de algoritmos encontrados em cada estudo

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	20	21	22
<i>Part-Of-Speech Tagger</i>	✓	✓	✓		✓			✓	✓	✓		✓	✓	✓	✓	✓	✓	✓		✓	✓	✓	
Sinônimos e desambiguação de termos			✓								✓									✓			
Remoção de <i>stopwords</i> e símbolos	✓								✓		✓						✓	✓					
<i>Lemmatization and Stemming</i>			✓											✓	✓		✓	✓					
Expressões regulares		✓																					
Sujeito-Verbo-Objeto			✓		✓																		
Resolução de Anáfora	✓											✓											
<i>Term Frequency (TF-IDF)</i>		✓						✓	✓											✓	✓		
<i>K-nearest-Neighbours (KNN)</i>				✓								✓											
<i>Pearson Criterion</i>												✓											
<i>Cosine Distance</i>												✓											
<i>Latent Semantic Analysis</i>												✓											

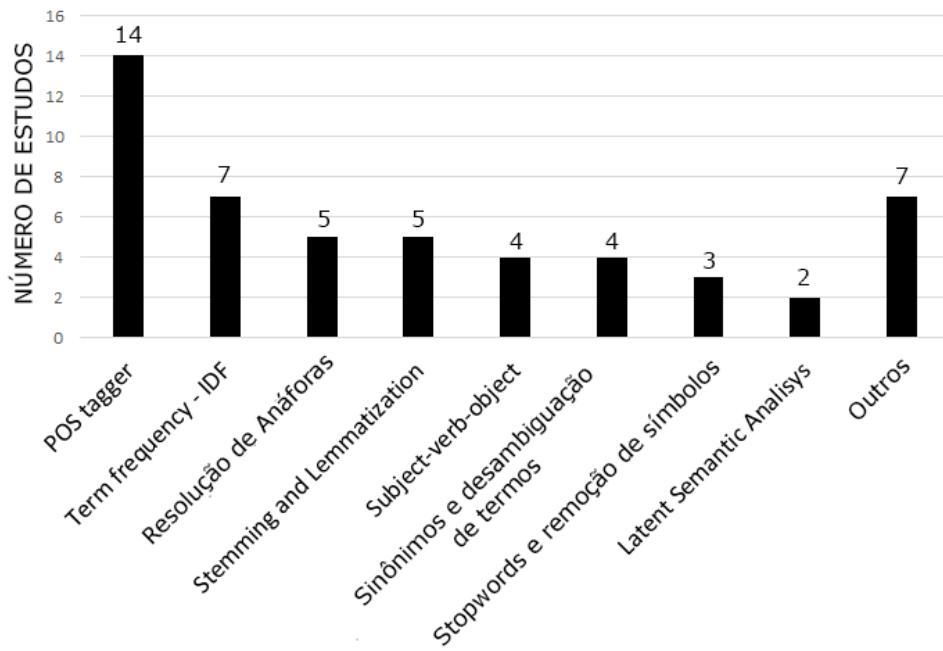


Figura 11: Algoritmos para construção de MCs

4.3.1 FRAMEWORK PARA CRIAÇÃO DE MCS BASEADO EM PLN

A combinação de técnicas específicas para resolver problemas pontuais na extração de MCs é bastante utilizada. Esta prática gera abordagens que utilizam um conjunto de técnicas cada vez mais híbridas. Buscando sumarizar esta prática e reunir informações para adotar a melhor estratégia para extração dos MCs, um *framework* foi desenvolvido para apresentar como as ferramentas e algoritmos para PLN podem ser empregadas na construção de MCs. A Figura 12 apresenta a estrutura desse *framework*. O *framework* foi gerado a partir dos resultados do mapeamento, especialmente os resultados das questões QP4, QP5 e QP6.

O *framework* visa orientar novas implementações associadas à construção de MCs usando técnicas PLN. A Figura 12 apresenta as principais técnicas que foram abordados nos 23 artigos selecionados no mapeamento, bem como a forma com que esses conceitos podem ser distribuídos em quatro quadrantes de acordo com duas dimensões: métodos linguísticos e estatísticos; seleção e associação de conceitos.

Dois quadrantes do *framework* estão relacionados aos algoritmos e ferramentas encontrados nos 23 estudos selecionados no MS. O *framework* apresenta os algoritmos e ferramentas de acordo com a finalidade para a qual são utilizadas, isto é, considerando se determinado algoritmo é utilizado para **associação de conceitos** ou de **seleção de conceitos**. A seleção de conceitos visa obter de um texto as palavras que melhor representam o seu domínio. Por outro

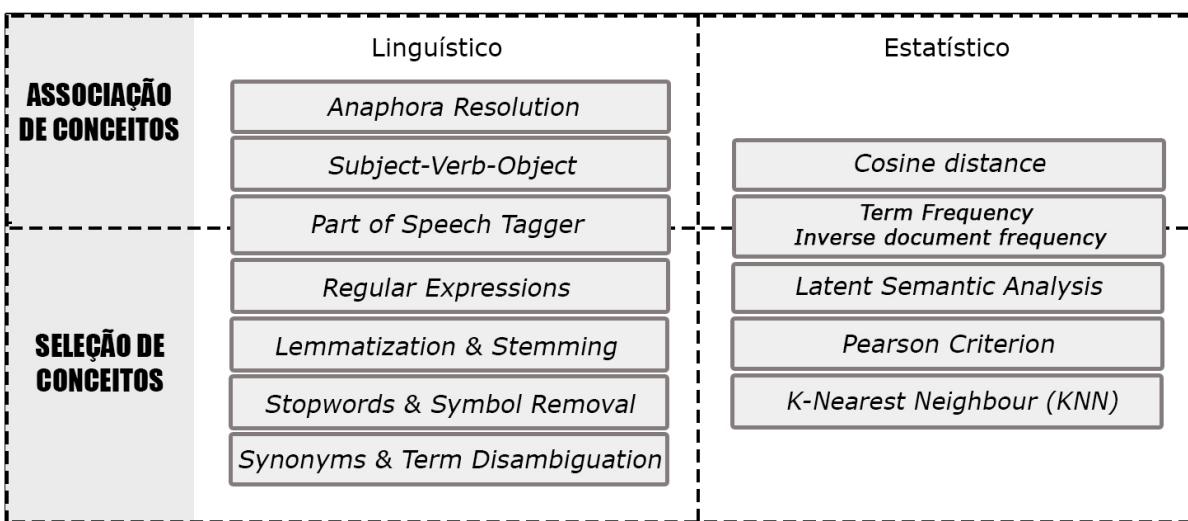


Figura 12: Framework para construção dos MCs baseados em PLN

lado, a associação de conceitos visa detectar possíveis conexões entre conceitos previamente selecionados.

Além disso, no *framework* os algoritmos e as ferramentas são apresentados de acordo com as técnicas em que são enquadradas, sendo linguísticas ou estatísticas. Essas técnicas fazem parte dos outros dois quadrantes da estrutura. A abordagem linguística usa as regras do idioma para selecionar conceitos e extraír relacionamentos. A técnica estatística pressupõe que a frequência de certos termos são indicadores de sua importância em um domínio específico. Além disso, as técnicas estatísticas ainda podem propor relacionamentos com base na frequência dos termos na mesma frase.

Com base na estrutura definida, um grafo de soluções foi elaborado (Figura 13). O objetivo do grafo é apresentar diferentes caminhos de solução para construir MC usando PLN. Essas soluções foram detectadas nos 23 estudos selecionados no mapeamento. Cada nó do grafo representa um algoritmo ou ferramenta usada nos estudos selecionados e sua frequência. As setas representam a direção que foi seguida ao usar um conjunto de algoritmos ou ferramentas, bem como as diferentes técnicas.

O grafo de solução permite iniciar a construção do MC a partir de dois nós distintos, representados pelos algoritmos POS Tagger e *TF-IDF*. O primeiro nó é representado por uma técnica linguística, neste caso, o algoritmo de POS tagger. O POS tagger foi escolhido como nó raiz visto que foi a ferramenta mais citada pelos estudos considerados linguísticos. O segundo nó raiz é representado por uma técnica estatística. O algoritmo estatístico *TF-IDF* foi o mais mencionado dentre os estudos classificados como estatístico.

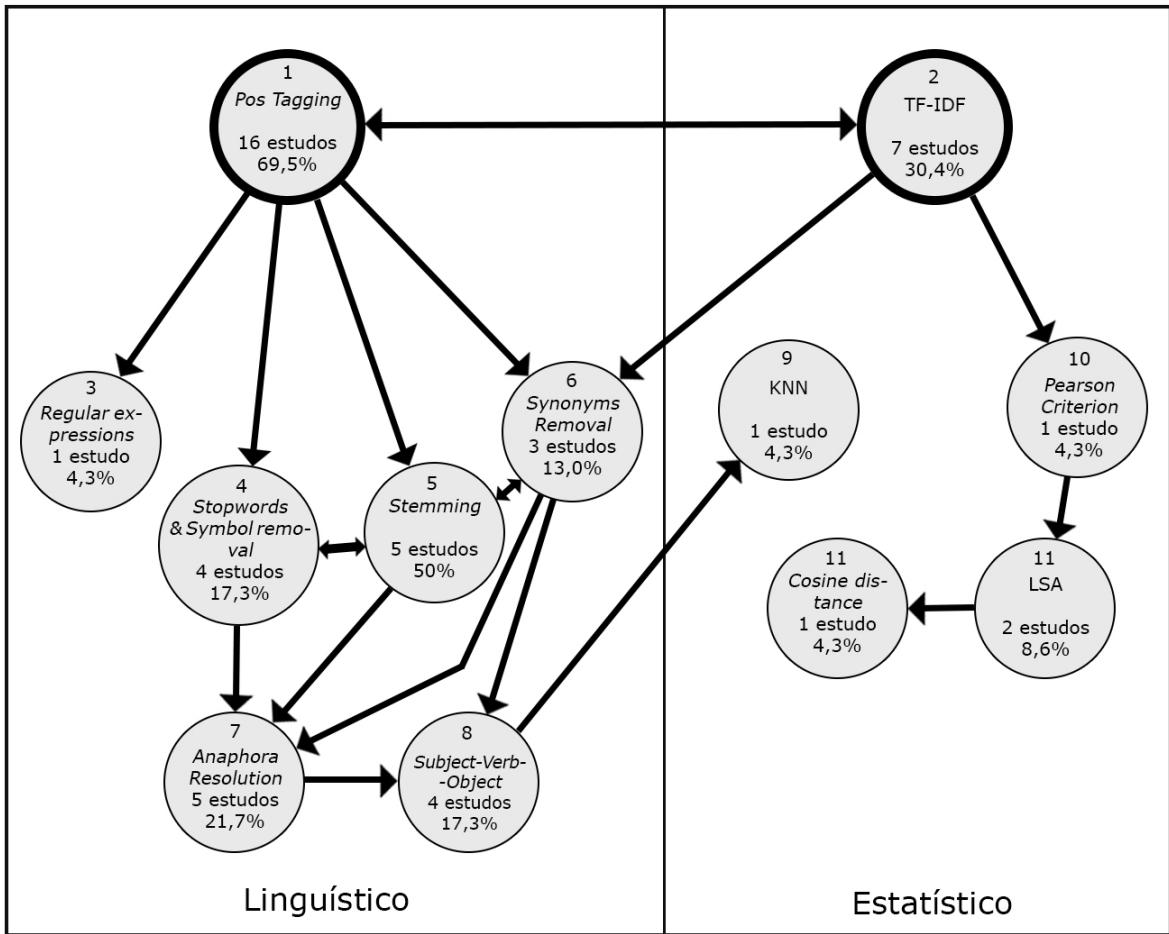


Figura 13: Grafo de soluções

O grafo apresenta caminhos possíveis considerados como uma solução viável para gerar MCs. Observe que, no grafo de solução, cada nó foi enumerado para facilitar o exemplo de possíveis caminhos de soluções. Para exemplificar tais soluções, são apresentados abaixo três exemplos reais considerando as diferentes técnicas de PLN.

Solução estatística: representa uma solução de algoritmos e ferramentas considerando apenas técnicas estatísticas para alcançar a construção do MC a partir de PLN. Um exemplo de caminho estatístico a seguir no grafo de solução é considerar a sequência de nós: [2, 10, 11, 12].

Solução linguística: representa uma solução de algoritmos e ferramentas considerando apenas técnicas linguísticas para alcançar a construção de MC da PLN. Um exemplo de caminho linguístico a seguir no grafo de solução é considerar a sequência de nós: [1, 3].

Solução híbrida: representa uma solução que combina algoritmos e ferramentas que consideram técnicas linguísticas e estatísticas para alcançar a construção de MCs. Um exemplo

de caminho híbrido a seguir no grafo de solução é considerar a sequência de nós: [1, 6, 8, 9].

4.4 CONSIDERAÇÕES FINAIS

De acordo com o MS realizado, a criação automática de MCs a partir de texto é um desafio de pesquisa valioso. Conforme apresentado neste estudo, embora o uso de PLN para construir MCs seja recente, essa abordagem tem sido bastante investigada nos últimos anos.

Após a realização deste estudo é possível estabelecer que a construção automática de MCs foi apresentada de forma positiva. Algumas das principais vantagens são apresentadas nos estudos, tais como: estimula especialistas a explorar novos métodos de resolução de problemas, de modo a aumentar a qualidade de suas soluções; facilitar a visualização e organizar o fluxo de ideias para estruturar conceitos-chave, bem como melhorar sua compreensão do texto; auxiliar no processo de ensino; e geração de MCs com rapidez, eficácia e integridade.

Considerando todos os aspectos dos estudos selecionados neste mapeamento, foi desenvolvido um framework para apresentar as ferramentas PLN e algoritmos poderiam ser usados em conjunto para a construção de MCs. Além disso, um grafo de solução foi elaborado para apresentar diferentes caminhos de solução para construir MCs usando PLN. Tanto o *framework* quanto o grafo podem ser usados para enumerar todas as soluções ou abordagens possíveis considerando as técnicas, algoritmos e ferramentas utilizadas nos 23 estudos selecionados pelo mapeamento. Essa enumeração servirá principalmente para orientar os pesquisadores da área a definir a melhor abordagem para usar para construir um MC, bem como para os futuros trabalhos na área.

5 ABORDAGEM PARA GERAÇÃO DE MCS A PARTIR DE PLN

Considerando os estudos realizados sobre MCs na área da Ciência da Computação que foram detalhados nos Capítulos 2, 3 e 4. Nesse capítulo é apresentada uma abordagem para construção de MCs a partir de PLN. A abordagem apresentada neste trabalho é baseada na estrutura proposta por Takemiya e Felizardo (2016) para representação de um artigo por meio de resumos gráficos. Este modelo foi apresentado na Seção 2.2.

Usando o modelo apresentado, o objetivo da iniciativa proposta é extrair conceitos e anexá-los no terceiro nível do modelo. A iniciativa foi dividida em 4 passos: (I) Entrada - Alimentar o sistema com resumos estruturados; (II) Pipeline 1 - extração de conceitos e relacionamentos; (III) Pipeline 2 - classificação; e (IV) Saída - Sumarização dos MCs baseados nas diretrizes propostas e resultados da apresentação. Estes passos são apresentados na Figura 14.

5.1 ENTRADA - RESUMOS ESTRUTURADOS

A entrada do sistema é baseada em resumos estruturados de artigos científicos. Estes resumos são divididos em cinco partes: contexto, objetivos, métodos, resultados e conclusões. A iniciativa apresentada assume que a entrada já está no formato estruturado.

Para auxiliar os pesquisadores nesta tarefa, iniciativas como a de Dayrell et al. (2012) podem ser utilizadas. Neste trabalho os autores propõem a separação automática de cinco partes dos resumos usando PLN. A ferramenta denominada MAZEA¹ (*Multi-label Argumentative Zoning for English Abstracts*) é um classificador que utiliza aprendizado de máquina para identificar automaticamente movimentos retóricos em resumos escritos na língua inglesa. Esta iniciativa pode classificar uma sentença em: (I) contexto, (II) lacuna, (III) propósito, (IV) método, (V) resultado e (VI) conclusão. Na sua versão inicial, o MAZEA concentrou-se em dois grandes campos: ciências físicas e engenharia e ciências da vida e da saúde. O MAZEA tem sido utilizado com sucesso em cursos de escrita científica em inglês como recurso pedagógico.

¹<http://www.nilc.icmc.usp.br/mazea-web/>

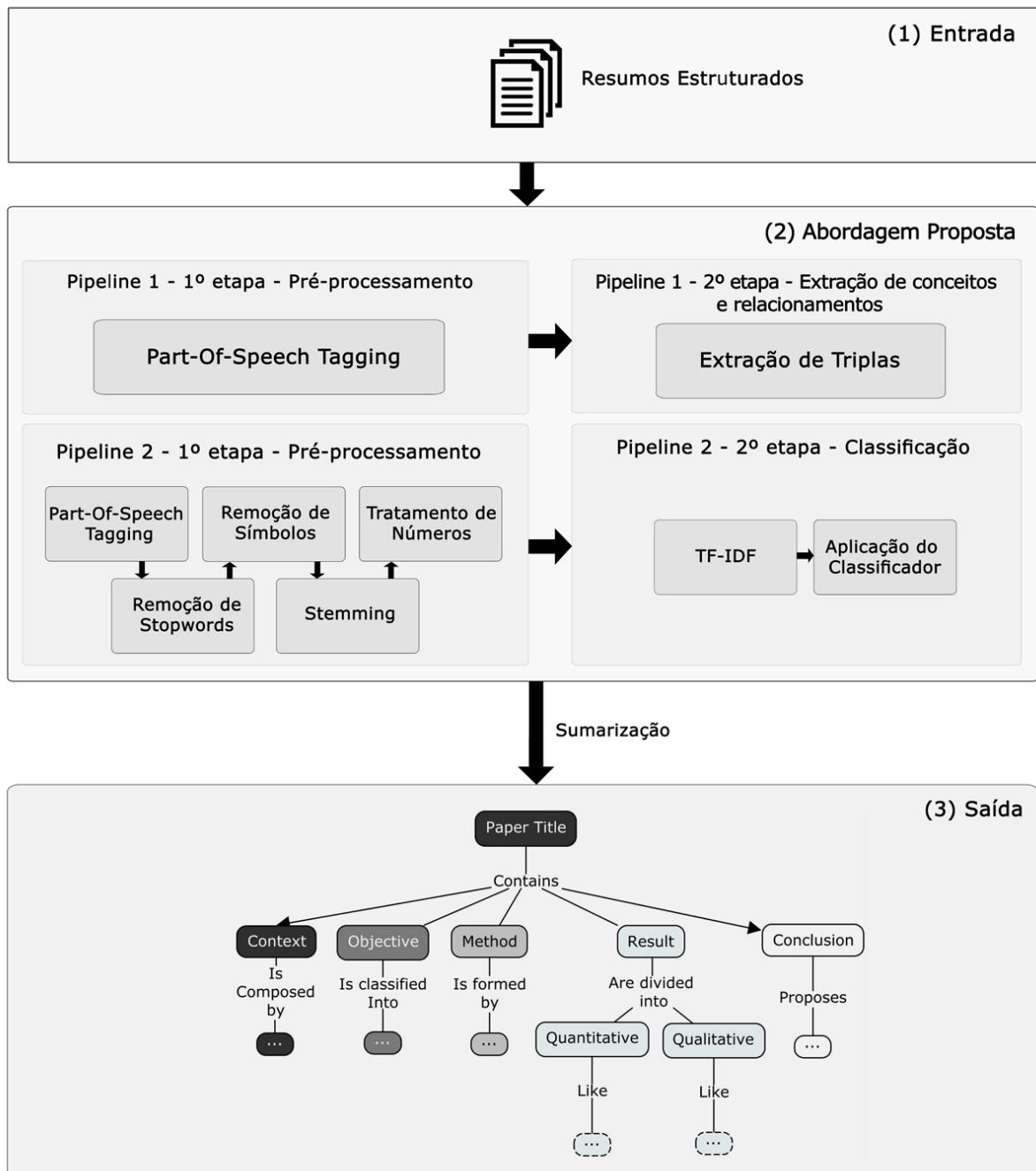


Figura 14: Iniciativa de extração e sumarização de MCs

5.2 PIPELINE 1 - EXTRAÇÃO DE CONCEITOS E RELACIONAMENTOS

Para extrair um MC que representa uma ideia completa é necessário identificar os conceitos e relacionamentos que serão utilizados. Na maioria dos textos é comum a presença de sentenças formadas por Sujeito-Verbo-Objeto. Esta estrutura também pode ser chamada de Tripla. Por exemplo, na sentença “João tem uma bicicleta” é possível perceber que “João” é o sujeito, “é” é um verbo e “bicicleta” é um objeto. Para identificar sentenças que apresentam esta estrutura, é necessário analisar o resultado de um algoritmo de *part-of-speech* e selecionar trechos que possuam este padrão explícito.

O principal problema em extrair conceitos e relacionamentos é a variação da língua. Em um texto longo, as sentenças não aparecem nesta estrutura. Para resolver este problema técnicas complementares são usadas. Por exemplo, a resolução de anáforas é usada por Qasim et al. (2013) e Aguiar et al. (2015) para detectar relacionamentos mais complexos entre sentenças que usam pronomes para representar um sujeito anteriormente citado. Por exemplo: “**Chris** nasceu em uma cidade pequena. **Ele** usa uma bicicleta para ir ao trabalho”. Neste exemplo o sujeito “Chris” foi substituído por um pronome “Ele”. Esta estrutura compõe uma anáfora.

O objetivo do primeiro pipeline é analisar um texto plano e extrair possíveis conceitos e relacionamentos que fazem sentido. Para alcançar este objetivo foi utilizada uma ferramenta desenvolvida pela Universidade de Stanford em seu grupo de pesquisa de PLN. Esta ferramenta é chamada de *Information Extractor* (IE)².

A abordagem proposta por Angeli et al. (2015) possui duas etapas. Na primeira etapa um classificador é treinado para separar sentenças em partes menores. As sentenças são reduzidas ao máximo, porém buscando manter as informações de contexto necessárias para compreensão do leitor. O algoritmo de *part-of-speech* foi utilizado para classificar cada palavra de acordo com sua função morfológica. Os autores definiram um pequeno conjunto de 14 padrões para segmentar as sentenças já classificadas pelo algoritmo de *part-of-speech* em uma tripla. Na segunda etapa, para cada cláusula extraída são produzidos argumentos que podem ser utilizados em outras aplicações. Os autores do IE apresentaram na avaliação que o seu sistema tem um bom desempenho em um mundo real (ANGELI et al., 2015)

²<https://nlp.stanford.edu/software/openie.html>

5.3 PIPELINE 2 - CLASSIFICAÇÃO

O propósito deste pipeline é treinar um classificador capaz de distinguir quando uma amostra pertence a uma determinada classe. Parte das classes analisadas foram definidas por Takemiya e Felizardo (2016) que podemos observar na Figura 3. Para criar o classificador foi definido um pipeline. A seguir, são apresentados o processo para criar a base de dados, incluindo o pre-processamento e como o classificador foi treinado.

O primeiro passo foi definir o pré-processamento. Este processo foi completado através: (1) remoção de *stopwords* (pronomes, conjunções, interjeições); (2) remoção de símbolos indesejados por meio de uma lista de símbolos; (3) aplicação da redução de palavras aos seus radicais buscando reduzir variações do mesmo termo; e (4) substituir números encontrados por um único marcador, esta conversão ajudará na análise e treinamento do classificador.

No segundo passo foi utilizado o algoritmo de *Term-Frequency* para extrair a frequência de um termo em um documento. Uma variação específica deste algoritmo foi utilizada, para que ele considerasse também a frequência do termo analisado em um conjunto de dados. Nessa fase o objetivo é construir um Saco de Palavras(“*Bag of Words*”) e aplicar o J48 para análise da frequência das palavras. Foi utilizado o classificador J48 para construção de uma árvore de decisão que armazena um modelo capaz de avaliar uma nova amostra.

O J48 é uma extensão de um algoritmo mais antigo denominado ID3. Ele é um classificador que faz parte do conjunto de algoritmos de aprendizado de máquina que realizam o aprendizado supervisionado. Este tipo de aprendizado analisa exemplos em uma base de dados e tenta predizer a classificação de uma nova amostra com base no modelo gerado ao treinar o classificador. Na ferramenta de mineração de dados WEKA³, o J48 é uma implementação Java de código aberto do algoritmo C4.5 (extensão do ID3). A ferramenta fornece uma série de opções associadas à poda de árvores e exibe uma avaliação da eficiência do classificador, mostrando números como a precisão, acurácia, *f-measure*, entre outras. Esta avaliação torna a interpretação dos resultados bastante amigável.

5.4 SAÍDA - SUMARIZAÇÃO DOS MCS

O passo final é agrupar as diretrizes com os conceitos e relacionamentos. Usando o classificador treinado no Pipeline 2, é possível selecionar o nó para anexar os conceitos extraídos no Pipeline 1. Após esta conexão é feito o mapa pronto para o usuário avaliar.

³<https://www.cs.waikato.ac.nz/ml/weka/>

5.5 CONSIDERAÇÕES FINAIS

Nesta seção foi apresentada uma abordagem que utiliza como entrada um resumo estruturado e passa por dois *pipelines* para gerar um MC. No primeiro *pipeline* são extraídos conceitos e relacionamentos que serão utilizados no MC. Já no segundo *pipeline*, um classificador treinado é utilizado para classificar cada sentença e relacioná-las a um conceito proposto pelo modelo de (FELIZARDO et al., 2017). Por fim, o MC é sumarizado conectando os resultados do *pipeline* 1 aos resultados indicados pelo *pipeline* 2.

No próximo capítulo são apresentadas as avaliações da aplicação desta abordagem proposta e os resultados obtidos.

6 VALIDAÇÃO DA ABORDAGEM

Para validar a abordagem apresentada no Capítulo 5 foram utilizadas duas estratégias: (i) avaliação da eficiência do classificador; e (ii) avaliação por meio de um experimento controlado (experimento piloto) considerando a opinião de alunos de mestrado da linha de ES.

Nas avaliações apresentadas nesse capítulo foi considerado o elemento “Resultado” do resumo estruturado. De acordo com Takemiya e Felizardo (2016), o elemento “*Resultado*” do resumo gráfico pode ser refinado em dois conceitos opcionais: *Quantitativo* e *Qualitativo*. Os conceitos opcionais fazem parte do nível 2 da estrutura de representação do resumo gráfico definido pelas autoras. Posteriormente, o nível 3 contém conceitos chamados de *conceitos variáveis*, que podem ser inseridos (ou não) pelo usuário na instanciação do modelo. O foco das avaliações que foram conduzidas estão relacionadas à identificação de qual conceito opcional o resultado do artigo está relacionado (*Quantitativo* ou *Qualitativo*), bem como a identificação de conceitos variáveis e criação de relacionamentos do nível 3 (*links*).

6.1 AVALIAÇÃO DA EFICIÊNCIA DO CLASSIFICADOR

Para analisar a eficiência do classificador, foram selecionadas 497 estudos da área de teste de software e as sentenças pertencentes ao elemento “*Resultado*” foram manualmente classificadas pelos envolvidos **neste** projeto de mestrado em *Quantitativo* ou *Qualitativo*. A seguir, o conjunto de dados foi submetido a técnicas de pré-processamento. A primeira técnica utilizada foi a *stemming*. Essa técnica foi aplicada para reduzir o número de variações de uma palavra extraíndo apenas o seu radical. O conjunto original sem a aplicação do *stemmer* possuía 2.089 palavras. Quando o *stemmer* foi aplicado esse número foi reduzido para 1.481 palavras. Outra técnica aplicada foi a remoção de palavras que não possuem significado expressivo ao texto, chamadas de *stopwords*, por exemplo, artigos e preposições.

Ainda para realização do pré-processamento, foi utilizada a conversão dos números presentes em cada sentença em um termo genérico. Esse termo evidencia apenas a presença de números na sentença. Esse processo foi adotado visto que durante a avaliação das árvores

de decisão geradas pelo classificador, percebeu-se que a presença de números é relevante para a classificação. Uma justificativa é a que estudos quantitativos normalmente mostram os resultados usando porcentagens e técnicas estatísticas. A árvore de decisão (Figura 15) também evidencia a importância dos números para a classificação considerando que os números estão presentes no *token* raiz da árvore.¹

```

number#999 <= 0
|   models <= 0
|   |   evaluation <= 0
|   |   |   more <= 0: Qualitative (328.0/33.0)
|   |   |   more > 0
|   |   |       result <= 0
|   |   |       |       source <= 0.2
|   |   |       |           analysis <= 0.1: Qualitative (21.0/2.0)
|   |   |       |           analysis > 0.1: Quantitative (3.0/1.0)
|   |   |       |       source > 0.2: Quantitative (2.0)
|   |   |       result > 0: Quantitative (2.0)
|   |   evaluation > 0
|   |       detection <= 0.1
|   |       |       evaluation <= 0.4: Quantitative (6.0/1.0)
|   |       |       evaluation > 0.4: Qualitative (4.0)
|   |       detection > 0.1: Qualitative (7.0)
models > 0
|   crosscompany <= 0
|   |   when <= 0: Qualitative (14.0)
|   |   when > 0: Quantitative (2.0)
|   crosscompany > 0: Quantitative (5.0)
number#999 > 0
|   number#999 <= 0.1
|   |   result <= 0.1
|   |   |   model <= 0: Quantitative (21.0/3.0)
|   |   |   model > 0: Qualitative (3.0)
|   |   result > 0.1: Qualitative (2.0)
number#999 > 0.1
|   quality <= 0.1: Quantitative (74.0/4.0)
|   quality > 0.1
|       |   quality <= 0.2: Qualitative (2.0)
|       |   quality > 0.2: Quantitative (2.0)

```

Figura 15: Árvore de decisão gerada pelo classificador que não utiliza *stemmer*

Buscando reduzir a dimensionalidade da base de dados foram aplicadas técnicas como: *Latent Semantic Analysis (LSA)*, *Principal Component Analysis (PCA)* e *Correlation-based Feature Subset Selection (CFSsubsetEval)*. Essas técnicas buscam selecionar os termos que possuem maior representatividade para criação do modelo e desprezar termos específicos ou genéricos que não auxiliam na criação da árvore de decisão (DUMAIS et al., 1988). Assim como o *LSA*, o *PCA* e o *CFSsubsetEval* são técnicas que também podem ser utilizadas para redução de dimensionalidade.

¹A base de dados utilizada está disponível no link: <https://goo.gl/kVsmeX>

Para avaliar qual técnica apresenta melhores resultados em reduzir a dimensionalidade e atinge maior precisão na classificação, todas foram aplicadas utilizadas e os resultados são comparados nas Figuras 16 e 17.

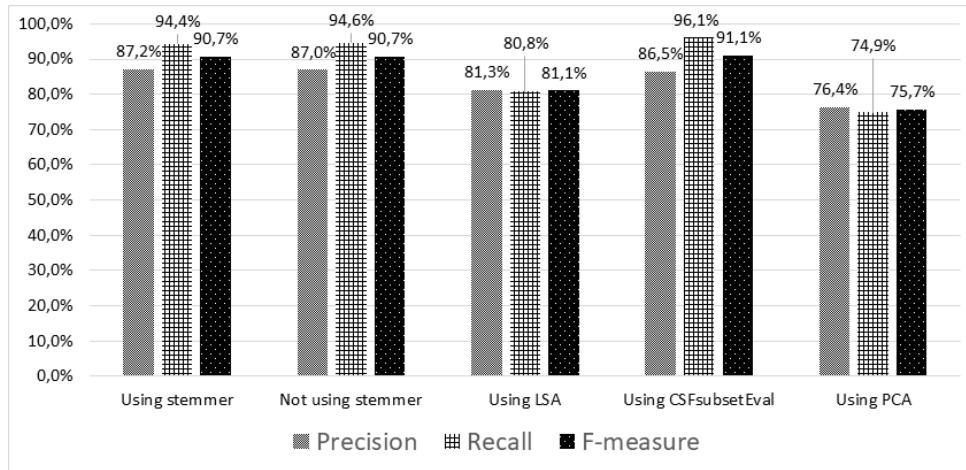


Figura 16: Avaliação do classificador de resumos considerando a classe qualitativa

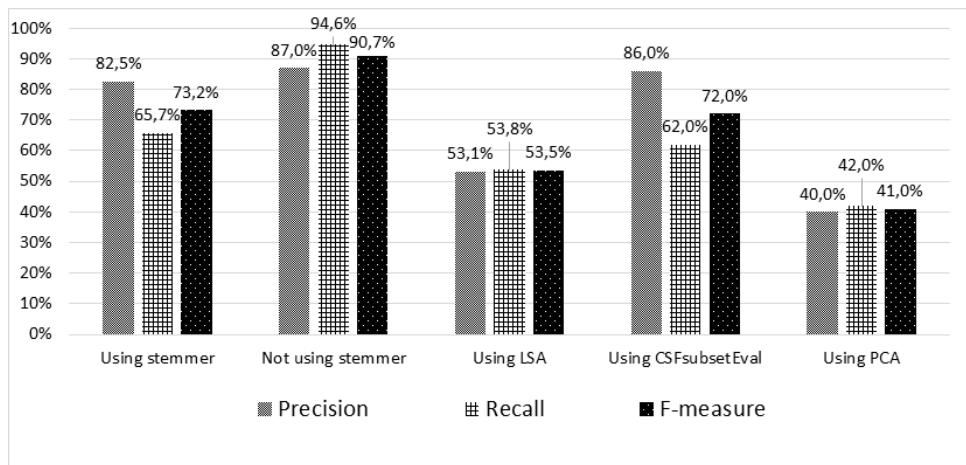


Figura 17: Avaliação do classificador de resumos considerando a classe quantitativa

Quando o *stemmer* foi utilizado para classificar as sentenças quantitativas alcançou-se a Precisão de 82,5%, um *Recall* de 65,7% e um *F-measure* de 73,2%. O classificador alcançou uma Precisão de 87,2%, um *Recall* de 94,4% e uma *F-measure* de 90,7% para sentenças qualitativas.

A não aplicação do *stemmer* para classificar as sentenças quantitativas gerou uma Precisão de 87,5%, *Recall* de 94,6% e *F-measure* de 90,7% . Já em relação às sentenças qualitativas, o classificador alcançou uma precisão de 87%, um *Recall* de 94,6% e um *F-measure* de 90,7%.

Em relação à técnica de *LSA* foi alcançada uma Precisão de 53,1% , um *Recall* de

53,8% e um *F-measure* de 53,5% ao classificar sentenças quantitativas. Já em sentenças qualitativas alcançou-se a Precisão de 81,3%, *Recall* de 80,8% e um *F-measure* de 81,1%.

Quando aplicada a técnica de *PCA* foi alcançada uma precisão de 40%, um *Recall* de 42% e um *F-measure* de 41% ao classificar sentenças quantitativas. Em sentenças qualitativas a Precisão foi de 76,4%, *Recall* de 74,9% e *F-measure* de 75,7%.

Já em relação à técnica de *CSFsubsetEval* foram alcançados 86% de precisão, 62% de *Recall* e 72% de *F-measure* quando avaliadas sentenças quantitativas. Em sentenças qualitativas foi alcançada a Precisão de 86,5%, *Recall* de 96,1% e *F-measure* de 91,1%.

Considerando os resultados da avaliação das técnicas de redução de dimensionalidade aplicadas foi possível perceber que o *dataset* sem a utilização de nenhuma técnica teve uma melhor performance.

A Figura 15 apresenta a árvore de decisão gerada pelo algoritmo J48 quando não utilizado o *stemmer* e nenhuma técnica de redução de dimensionalidade.

6.2 VALIDAÇÃO DOS MCS CONSTRUÍDOS AUTOMATICAMENTE

Nessa seção será demonstrado como foi feita a criação de um MC utilizando a abordagem proposta no Capítulo 5. A seguir são apresentados os detalhes do experimento conduzido para avaliar os MCs gerados.

6.2.1 CONSTRUÍNDΟ UM MC AUTOMATICAMENTE

Nesta seção um MC será construído utilizando a abordagem descrita no Capítulo 5. Para demonstrar o funcionamento da abordagem proposta foi escolhido o estudo de Santos et al. (2017). Neste estudo o autor propõe a condução de um MS na área de ciência da computação para compreender a utilização dos MCs nesta área.

O processo de construção automática do MC será apresentado 4 passos, são eles: (I) Entrada de dados, (II) Pipeline I - Extração de triplas, (III) Pipeline II - Classificação, e (IV) sumarização.

(I) Entrada de dados

Os dados utilizados como entrada da abordagem fazem parte de resumo estruturado. Nesse resumo, os elementos **contexto, objetivos, métodos, resultados e conclusões**, foram

previamente identificados pelo autor. Observe a seguir o resumo estruturado extraído de (SANTOS et al., 2017).

Context: Concept Maps (CMs) enable the creation of a schematic representation of a domain knowledge. For this reason, CMs have been applied in different research areas, including Computer Science. **Objective:** the objective of this paper is to present the results of a systematic mapping study conducted to collect and evaluate existing research on CMs initiatives in Computer Science. **Method:** the mapping study was performed by searching five electronic databases. We also performed backward snowballing and manual search to find publications of researchers and research groups that accomplished these studies. **Results:** from the mapping study, we identified 108 studies addressing CMs initiatives in different subareas of Computer Science that were reviewed to extract relevant information to answer a set of research questions. The mapping shows an increasing interest in the topic in recent years and it has been extensively investigated due to support in teaching and learning. **Conclusions:** based on our results we conclude that the use of CMs as an educational tool has been widely accepted in Computer Science.

Para esta demonstração foi utilizado apenas o elemento “Resultado” do resumo estruturado.

(II) Pipeline I - extração de triplas

Ao submeter os resultados do resumo estruturado ao pipeline I, as sentenças são quebradas em sentenças menores. Observe no exemplo a seguir como é possível quebrar o texto do elemento “resultado” em sentenças menores:

1. *from the mapping study, we identified 108 studies*
2. *addressing CMs initiatives in different subareas of Computer Science that were reviewed to extract relevant information to answer a set of research questions.*
3. *The mapping shows an increasing interest in the topic in recent years*
4. *and it has been extensively investigated due to support in teaching and learning.*

Em seguida, utilizando o algoritmo de *Part-of-speech Tagging* cada palavra é classificada de acordo com sua função na sentença. Observando o exemplo presente na Figura 18 a frase “We identified 108 studies” foi classificada de acordo pelo algoritmo de *POS-tagging*.

Observando as *Tags* atribuídas, as palavras encontramos o padrão “NP-VBD-NP” que configura um relacionamento válido de acordo com a abordagem de Angeli et al. (2015). Desta forma o *Information Extractor* busca por padrões similares que configuram um relacionamento. Os autores Angeli et al. (2015) definiram 14 padrões diferentes para identificar esses possíveis relacionamentos.

(NP) *we* (VBD) *identified* (NP) *108 studies*

Legenda:

(NP) → Substantivo próprio, singular
 (VBD) → Verbo, Pretérito

Figura 18: Exemplo de saída do algoritmo de *POS-Tagging*

Como resultado o *Information Extractor* retorna possíveis (C) conceitos, que são conectados por possíveis (R) relacionamentos:

- (C) *we* – (R) *identify* – (C) *108 study*
- (C) *mapping* – (R) *show* – (C) *increase interest in recent year*
- (C) *it* – (R) *have* – (C) *have extensively investigate*
- (C) *CMs initiative* – (R) *be in* – (C) *different subarea of computer science*

(III) Pipeline II - Classificação

Nessa etapa foi utilizado o classificador já treinado para predizer se uma sentença pertence ao elemento “Quantitativo” ou “Qualitativo”. A sentença “*We identify 108 studies*”, por exemplo, foi classificada como um elemento quantitativo. Por outro lado, a sentença ‘*The mapping shows an increasing interest in the topic in recent years*’ pertence aos resultados qualitativos do trabalho.

Utilizando as informações extraídas na aplicação do classificador é possível realizar a sumarização que é descrita a seguir.

(IV) Sumarização

Após a execução dos Pipelines I e II, a sumarização do MC é feita anexando as triplas extraídas (Pipeline I) ao modelo para estudos científicos proposta por (FELIZARDO et al., 2017) utilizando as previsões obtidas do classificador (Pipeline II). O resultado é apresentado na Figura 19.

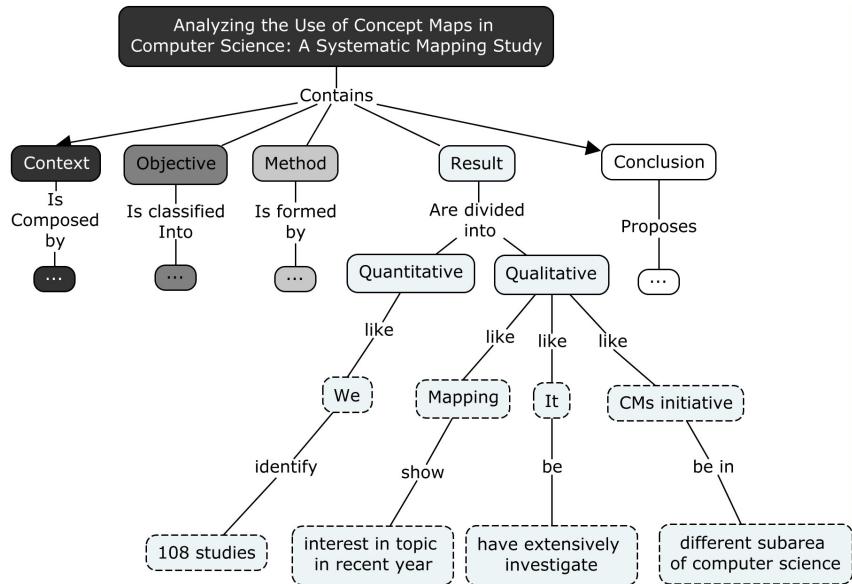


Figura 19: Exemplo criado pela abordagem proposta

A seguir serão apresentados os detalhes do estudo piloto para avaliação da iniciativa.

6.2.2 CONDUÇÃO DO ESTUDO EXPERIMENTAL

O experimento piloto visou investigar a qualidade dos MCs gerados pela iniciativa proposta. Os MCs foram avaliados considerando a representatividade dos conceitos, a validade das ideias expressas pelos relacionamentos e o nível de cobertura.

As seguintes questões de pesquisa e hipóteses foram definidas:

Questão de pesquisa 1: Os conceitos apresentados nos MCs representam ideias e informações relacionadas ao elemento “*Result*” do resumo correspondente?

- Hipótese nula (H0): *Os conceitos apresentados nos MCs não representam ideias e informações relacionadas ao elemento “Result” do resumo correspondente.*
- Hipótese alternativa (H1): *Os conceitos apresentados nos MCs representam ideias e informações relacionadas ao elemento “Result” do resumo correspondente.*
- Hipótese alternativa (H2): *Os conceitos apresentados nos MCs representam parcialmente as ideias e informações relacionadas ao elemento “Result” do resumo correspondente.*

Questão de pesquisa 2: Os relacionamentos conectam adequadamente os conceitos relacionados ao elemento “*Result*” do resumo correspondente?

- Hipótese nula (H0): *Os relacionamentos não conectam adequadamente os conceitos relacionados ao elemento “Result” do resumo correspondente.*
- Hipótese alternativa (H1): *Os relacionamentos conectam adequadamente os conceitos relacionados ao elemento “Result” do resumo correspondente.*
- Hipótese alternativa (H2): *Os relacionamentos apresentados representam ideias válidas, porém, representam parcialmente os conceitos relacionados ao elemento “Result” do resumo correspondente.*

Questão de pesquisa 3: O MC abrange todas as informações relacionadas ao elemento “*Result*” do resumo correspondente?

- Hipótese nula (H0): *O MC não abrange nenhuma informação relacionada ao elemento “Result” do resumo correspondente.*
- Hipótese alternativa (H1): *O MC abrange todas as informações relacionadas ao elemento “Result” do resumo correspondente.*
- Hipótese alternativa (H2): *O MC abrange parcialmente as informações relacionadas ao elemento “Result” do resumo correspondente.*

Os participantes selecionados devem ser representativos para o experimento. Sendo assim, para esse experimento foram escolhidos estudantes de mestrado habituados a realizar a leitura de artigos científicos da área da Ciência da Computação e capazes de compreender termos escritos na língua inglesa. O experimento foi organizado em duas etapas: treinamento e execução.

(1) Treinamento

Durante o treinamento os participantes receberam o documento presente no apêndice A. Para o treinamento foi utilizado o MC gerado do artigo (SANTOS et al., 2017).

O estudo presente no documento de treinamento continha o título do trabalho, o MC e o resumo com a seção a ser avaliada destacada no texto. Durante a atividade de avaliação os participantes deveriam ler o resumo do artigo, em especial, o resultado do artigo considerando o MC apresentado. No decorrer da etapa de treinamento os participantes esclareceram as dúvidas

e se prepararam para a etapa de execução do experimento. Para assegurar que a etapa de treinamento não interferisse na execução, os estudos utilizados na etapa de execução eram diferentes do utilizado no treinamento.

(2) Execução

Após a etapa de treinamento, foi conduzida a etapa de execução na qual os participantes receberam um documento contendo seis MCs e seus respectivos resumos para serem avaliados. Os estudos utilizados estão apresentados na Tabela 6.2.2.

ID	Título do estudo	Publicado em	Ano
1	Automatic detection of cross-browser incompatibilities using machine learning and screenshot similarity: student research abstract	SAC '17 Proceedings of the Symposium on Applied Computing	2017
2	Evaluating test suite characteristics, cost, and effectiveness of FSM-based testing methods	Information and Software Technology	2013
3	An empirical study to quantify the characteristics of Java programs that may influence symbolic execution from a unit testing perspective	Journal of Systems and Software	2016
4	Test oracles for simulink-like models	Journal Automated Software Engineering	2017
5	Knowledge management initiatives in software testing: A mapping study	Information and Software Technology	2015
6	The Use of Visual Text Mining to Support the Study Selection Activity in Systematic Literature Reviews: A Replication Study	Replication in Empirical Software Engineering Research	2013

Os seis estudos selecionados foram submetidos a iniciativa proposta no Capítulo 5. Foram gerados automaticamente seis MCs correspondentes aos dados presentes no resumo de cada estudo. Vale destacar que a avaliação foi feita apenas considerando o elemento “*Resultado*” no resumo dos estudos e todos eles utilizam o modelo proposto por (FELIZARDO et al., 2017).

(3) Divisão dos grupos

Visando diminuir o possível viés da pesquisa, foi solicitado aos autores dos estudos utilizados no experimento, que desenhassem o MC que representasse os resultados obtidos em seu trabalho, considerando apenas o que estava escrito no elemento “*Resultado*” do resumo do artigo. Ao preparar o material para execução do experimento e distribuí-los aos participantes, foram utilizadas as duas versões: (1) MCs gerados automaticamente (apêndice B); e (2) MCs elaborados pelos autores dos estudos (apêndice C). Essa divisão **não** foi comunicada aos participantes da pesquisa para reduzir o risco de interferência nos resultados.

6.2.3 RESULTADOS DO EXPERIMENTO

Os resultados apresentados nessa seção ainda são considerados preliminares e não foram avaliados estatisticamente devido a pequena amostragem de participantes. O estudo teve a participação de 32 estudantes da área de Ciência da Computação de 24 a 62 anos.

Para solucionar as questões de pesquisa, foram analisadas as respostas presentes no questionário entregue aos alunos. A seguir são apresentadas as avaliações feitas pelos estudantes utilizando a escala de *Likert* (LIKERT, 1932). Para cada questão o participante deveria marcar uma das opções: concorda fortemente, concorda, indeciso, discorda, discorda fortemente.

- Os conceitos apresentados no MC representam ideias e informações relacionadas ao elemento “*Resultado*” do resumo correspondente.
- Os relacionamentos (*Links*) utilizados conectam adequadamente os conceitos relacionados ao elemento “*Resultado*” do resumo correspondente.
- O MC apresentado abrange todas as informações relacionadas ao elemento “*Resultado*” do resumo correspondente.

QP1 - Avaliação dos conceitos utilizados

Ao aplicar o questionário, foi possível constatar que 79,33% dos participantes que classificaram os MCs gerados automaticamente, concordam fortemente ou concordam que os conceitos do MC representam as ideias e informações relacionadas ao elemento “*Resultado*” do resumo correspondente (Gráfico da Figura 20). Avaliando os MCs gerados pelos autores, em média 67,83% dos participantes concordaram fortemente ou concordaram que os conceitos extraídos representam ideias e informações relacionadas ao elemento “*Resultado*” do resumo correspondente. Esse resultado é apresentado na Figura 21. Com base nesses dados é possível inferir que os participantes consideram que a extração automática dos conceitos representa de forma satisfatória o elemento “*Resultado*” dos artigos selecionados para o experimento.

QP2 - Avaliação dos relacionamentos utilizados

Ao avaliar os conexões utilizadas nos MCs, foi possível constatar que em média 59% dos participantes concordam ou concordam fortemente que os MCs gerados automaticamente possuem relacionamentos (*links*) que conectam adequadamente os conceitos do elemento “*Resultado*” do resumo correspondente (ver o gráfico da Figura 22). Com relação aos MCs gerados pelos autores, em média 61% dos participantes concordaram fortemente ou concordaram que os MCs gerados automaticamente possuem relacionamentos que conectam adequadamente os

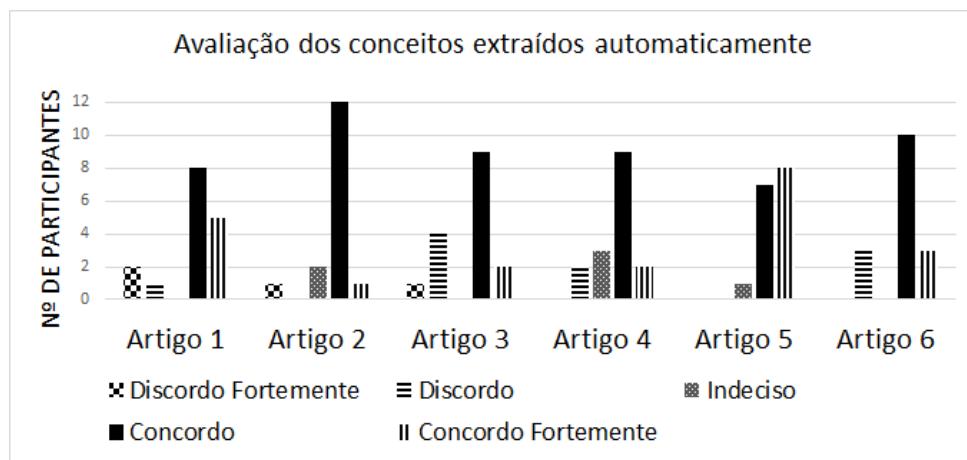


Figura 20: Avaliação dos conceitos apresentados nos MCs gerados automaticamente

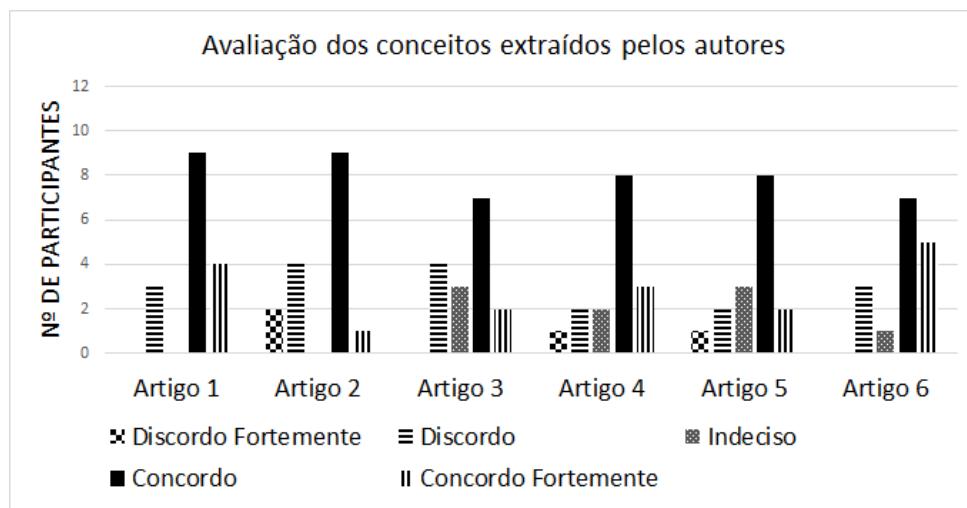


Figura 21: Avaliação dos conceitos apresentados nos MCs gerados pelos autores

conceitos do elemento “*Resultado*” do resumo correspondente. O gráfico da Figura 23 apresenta os resultados da avaliação referente aos relacionamentos do MC criados pelos autores dos estudos.

QP3 - Avaliação da abrangência dos MCs

Quando questionados sobre a abrangência dos MCs, em média 71% dos participantes do experimento, concordaram ou concordaram fortemente que os MCs gerados automaticamente abrangem todas as informações relacionadas ao elemento “*Resultado*” do resumo correspondente, conforme apresentado pelo gráfico da Figura 24. Já em relação à abrangência dos MCs gerados pelos autores, em média 67% dos participantes concordaram fortemente ou concordaram que os MCs gerados abrangem todas as informações relacionadas ao elemento “*Resultado*” do resumo correspondente (Figura 25).

Pela análise dos resultados foi possível perceber que a iniciativa proposta se destaca

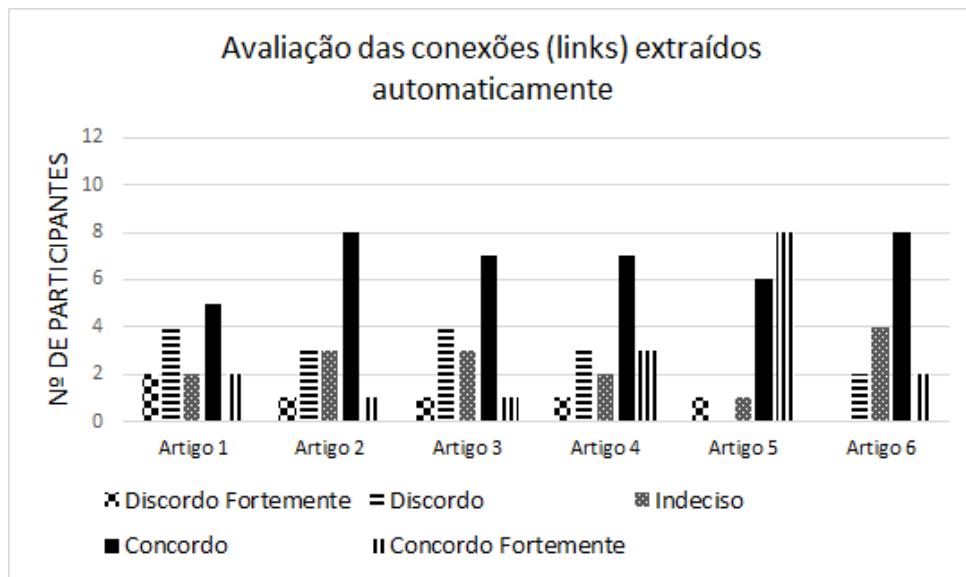


Figura 22: Avaliação dos relacionamento apresentados nos MCs gerados automaticamente

quando é necessário extrair conceitos e abranger todos os elementos apresentados no resumo. No entanto, pode-se verificar que ainda existe uma deficiência ao extrair relacionamentos representativos do texto. Observando os MCs gerados automaticamente (Apêndice B) é possível perceber que ao extrair os relacionamentos, muitos verbos são colocados no infinitivo e perdem sua conjugação. Isto acontece pois para a extração de triplas o algoritmo remove as possíveis variações causadas pela conjugação verbal. A falta de tempo verbal pode causar confusão no entendimento do MC e prejudicar sua leitura.

Ainda é importante ressaltar que existem resumos que não possuem as informações necessárias para representar corretamente os resultados ou qualquer outra seção, assim como já foi apresentado no Capítulo 1. Quando um resumo é mal escrito pelo autor, o algoritmo proposto pode apresentar dificuldades em localizar conceitos e relacionamentos. Esses movimentos que ainda não podem ser capturados pelo algoritmo quando não representados no MC pode torná-lo incompleto, ou seja, o MC não representa uma ideia completa e válida.

6.3 DISCUSSÃO

Algumas abordagens encontradas na literatura são semelhantes à iniciativa proposta por este trabalho. A abordagem de Bui et al. (2016) propõe um sistema de sumarização para aumentar a produtividade e reduzir erros no processo tradicional de extração de dados de estudos secundários. Com relação a este trabalho foram encontradas algumas diferenças importantes: (i) o padrão proposto pelos autores e a validação foi feita na área médica; (ii) a abordagem usa um único documento para treinar um modelo de regressão de aprendizado de máquina;

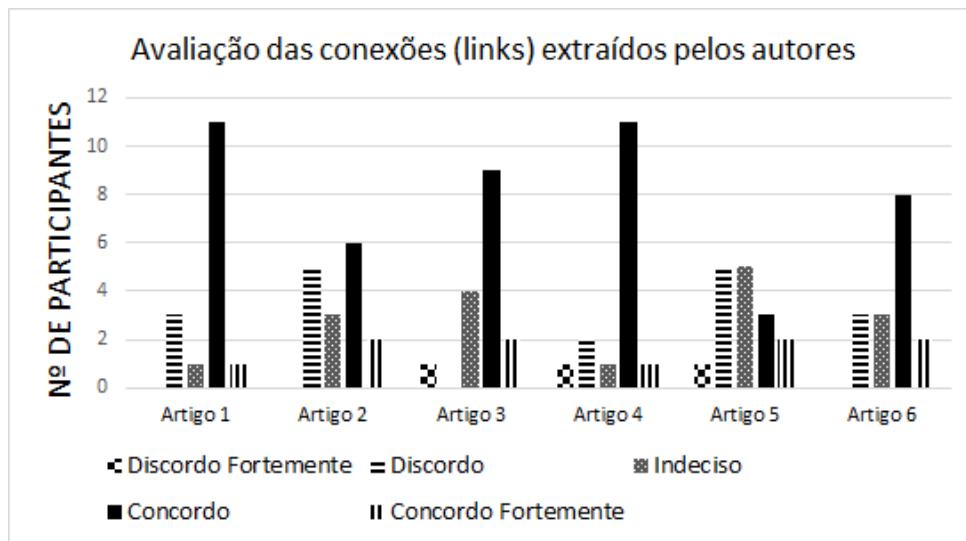


Figura 23: Avaliação dos relacionamento apresentados nos MCs gerados pelos autores

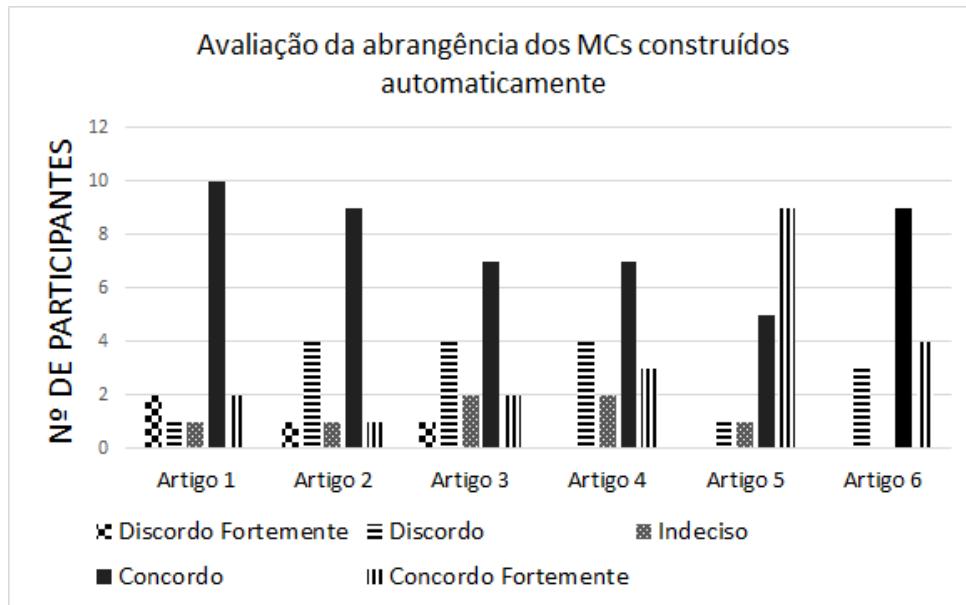


Figura 24: Avaliação da abrangência dos MCs gerados automaticamente
Autoria própria

(iii) a abordagem propõe auxiliar a fase de extração de dados e não a fase de seleção; e (iv) a abordagem não usa MCs para representação visual e auxílio da fase de extração de dados.

Outra abordagem importante foi proposta por Qasim et al. (2013). O objetivo de Qasim et al. (2013) foi ajudar a construir MCs de documentos de texto. Ao comparar as iniciativas propostas foram identificadas algumas semelhanças com a abordagem proposta neste trabalho, por exemplo, o uso da técnica de *part-of-speech*. Também foi identificado o uso de algumas variações de métricas, por exemplo, o VF-ICF é uma métrica modificada por Punuru e Chen (2012) e mostra que os verbos que ocorrem com mais frequência tendem a ser mais amplos e

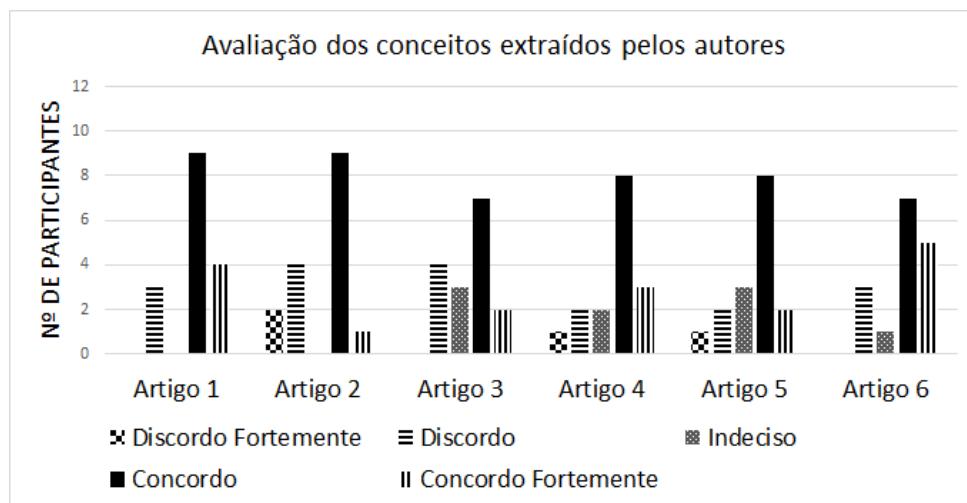


Figura 25: Avaliação da abrangência dos MCs gerados pelos autores
Autoria própria

não denotar relações semânticas importantes. Essa modificação não precisa de um conjunto de documentos diferentes, em vez disso, essa abordagem considera o estudo completo e não só o resumo. A técnica de *Affinity Propagation* também foi usada para agrupar termos relacionados. Por fim, a resolução de anáforas foi utilizada para extrair relacionamentos complexos que usam pronomes para representar os substantivos.

Aguiar e Cury (2016) propuseram um novo método automático para a geração de MCs. Uma importante diferença encontrada é o uso de uma abordagem semiautomática que o usuário escolhe domínio na entrada. Além disso, os autores mencionam que a resolução de anáforas ainda está longe de ser satisfatória. Essa técnica é citada como uma limitação do trabalho, visto que conceitos e relacionamentos relevantes aos domínios foram perdidos na extração.

A diferença mais importante entre a abordagem proposta neste trabalho e os demais estudos encontrados é o uso do modelo proposto por Takemiya e Felizardo (2016). O modelo ajuda os pesquisadores na coleta ou seleção de estudos na realização de um estudo secundário. Além disso, o modelo proposto fornece um padrão de estrutura que auxilia na leitura do MC.

Foram identificadas algumas ameaças quanto a validade do experimento piloto conduzido. Dentre elas destacam-se:

- Ameaça interna: a baixa quantidade de autores disponíveis para criação dos MCs utilizados no experimento limitou a quantidade de mapas utilizados no estudo.
- Ameaça na construção: o nível de experiência dos participantes no tema dos estudos analisados poderia afetar os resultados finais. Além disso, a língua nativa dos participantes

(português) era diferente do material dos estudos (inglês). No caso do experimento piloto conduzido a maioria dos participantes possuía um nível intermediário de leitura em inglês;

- Ameaça externa: é difícil generalizar os resultados considerando que a amostra usada foi pequena. Para minimizar essa ameaça o experimento deve ser replicado com uma amostragem maior de participantes.

7 CONCLUSÃO

O presente trabalho propôs uma abordagem para construção de MCs a partir de técnicas de Processamento de Linguagem Natural. Pretendeu-se com esta abordagem auxiliar autores a construir resumos visuais que podem auxiliar o processo de condução de estudos secundários em ES. Para alcançar este objetivo foram propostas seguintes atividades: (I) conduzir um mapeamento para identificar o estado da arte sobre o uso dos MCs na área de Ciência da Computação; (II) Conduzir um MS para identificar iniciativas de construção de MCs a partir de PLN; (III) Definir uma abordagem para geração automática de MCs com base em PLN utilizando as diretrizes propostas por Takemiya e Felizardo (2016); e (IV) Condução de avaliações com o objetivo de verificar a qualidade dos MCs gerados pela abordagem proposta.

7.1 PRINCIPAIS CONTRIBUIÇÕES E LIMITAÇÕES DA PESQUISA

Como contribuição, este estudo apresentou uma abordagem para construção de MC automaticamente utilizando técnicas de PLN. Além disso, os resultados de um experimento piloto que foi conduzido para avaliar a qualidade dos MCs gerados automaticamente. Os resultados desse estudo experimental indicam que os MCs gerados apresentaram conceitos e relacionamentos válidos, assim como uma boa abrangência. Ao utilizar a abordagem, os autores dos artigos podem receber automaticamente uma sugestão válida de como estruturar os MCs. Isto pode auxiliá-los na criação de MCs para representação de estudos científicos. Uma limitação desta pesquisa está relacionada a pequena amostra de participantes. Dessa forma, os resultados do experimento piloto não foi analisado estatisticamente.

7.2 TRABALHOS FUTUROS

Como trabalho futuro pretende-se aplicar a abordagem proposta aos outros elementos do resumo estruturado, como “Contexto”, “Objetivo”, “Método” e “Conclusão”. Outra oportunidade futura é construção de uma ferramenta para auxiliar os autores na geração de MCs.

7.3 PUBLICAÇÕES

Para divulgar os resultados deste projeto e cumprir os créditos solicitados pelo programa de mestrado foi realizada a publicação de um artigo científico abordando o uso de MCs na ciência da computação (Capítulo 3) na revista *Informatics in Education* no ano de 2017. Uma segunda publicação foi realizada no 33º ACM/SIGAPP Symposium On Applied Computing. A linha de pesquisa selecionada foi a de Representação de conhecimento e raciocínio (SAC 2018). O trabalho foi apresentado pelo aluno na cidade de Pau na França nos dias 10, 11 e 12 de Abril.

SANTOS, V., SOUZA, É. F., Felizardo, K. R., Vijaykumar, N. L. Analyzing the use of concept maps in computer science: a systematic mapping study. *Informatics and Education*, v. 16, n. 2, p. 257–288, 2017.

SANTOS, V., Concept Maps Construction Using Natural Language Processing to Support Studies Selection. *Informatics and Education*, 33rd ACM/SIGAPP Symposium On Applied Computing, Pau, França: 2018, In press.

REFERÊNCIAS

- AGUIAR, C. Z.; CURY, D. Automatic Construction of Concept Maps from texts. In: **4th International Conference on Concept Mapping**. Tallinn, Estonia: [s.n.], 2016. p. 1–6.
- AGUIAR, C. Z.; CURY, D.; GAVA, T. Um Estudo sobre Abordagens Tecnológicas para a Geração de Mapas Conceituais. In: **20th Congreso Internacinal de Informática Educativa**. Santiago, Chile: [s.n.], 2015. p. 136–146.
- ALIAS, M.; SURADI, Z. Concept mapping: A tool for creating a literature review. In: **3rd International Conference on Concept Mapping (CMC)**. Tallin, Estonia: [s.n.], 2008. p. 1–4.
- Álvarez Bermejo, J. A.; Codina Sánchez, A.; Belmonte Ureña, L. J. Application architecture to efficiently manage formal and informal m-learning. a case study to motivate computer engineering students. **DYNA**, v. 82, n. 190, p. 113–120, 2015.
- ANGELI, G.; PREMKUMAR, M. J. J.; MANNING, C. D. Leveraging linguistic structure for open domain information extraction. In: **53rd Anual meeting of Association for Computational Linguistics and 7th International Conference on Natural Language Processing**. Beijing, China: [s.n.], 2015. p. 344–354.
- ARANHA, C. N. **Uma Abordagem de Pré- Processamento Automático para Mineração de Textos em Português: Sob o Enfoque da Inteligência Computacional**. 1–149 p. Tese (Tese de Doutorado) — Pontifícia Universidade Católica, 2007.
- ARRUARTE, A. et al. Collaborative and Multilingual Approach to Learn Database Topics Using Concept Maps. **Scientific World Journal**, v. 2014, p. 1–8, 2014.
- ARRUARTE, A.; RUEDA, U.; ELORRIAGA, J. A. Organizing the learning resources related to the subject introduction to artificial intelligence through concept maps. In: **38th Annual Frontiers in Education Conference (FIE)**. Madrid, Espanha: [s.n.], 2008. p. 17–22.
- ATAPATTU, T.; FALKNER, K.; FALKNER, N. Educational Question Answering Motivated by Question-Specific Concept Maps. In: **Lecture Notes in Computer Science**. [S.l.: s.n.], 2015. v. 9112, p. 13–22.
- ATAPATTU, T.; FALKNER, K.; FALKNER, N. Task-Adapted Concept Map Scaffolding to Support Quizzes in an Online Environment. In: **ACM Conference on Innovation and Technology in Computer Science Education - ITiCSE**. Vilnius, Lithuania: [s.n.], 2015. p. 272–277.
- AUSUBEL, D. P. **The psychology of meaningful verbal learning**. [S.l.: s.n.], 1963.
- AUSUBEL, D. P.; NOVAK, J. D.; HANESIAN, H. **Educational psychology: A cognitive view**. Michigan, United States: Holt, Rinehart and Winston, 1968. 15–31 p.
- AUSUBEL, D. P.; NOVAK, J. D.; HANESIAN, H. **Educational Psychology: A Cognitive View**. 2. ed. New York: Warbel & Peck, 1978.

- BERGES, M.; HUBWIESER, P. Concept specification maps - displaying content structures. In: **18th ACM conference on Innovation and technology in computer science education -ITiCSE**. Canterbury, Inglaterra: [s.n.], 2013. p. 291–296.
- BERGES, M.; MÜHLING, A.; HUBWIESER, P. The gap between knowledge and ability. In: **12th Koli Calling International Conference on Computing Education Research**. Koli, Finland: [s.n.], 2012. p. 126–134.
- BLECIC, I.; CECCHINI, A.; TRUNFIO, G. A collaborative and collective concept mapping tool. **4th International Conference (CDVE)**, v. 4674 LNCS, p. 264–271, 2007.
- BRERETON, P. et al. Lessons from applying the systematic literature review process within the software engineering domain. **Journal of Systems and Software**, v. 80, n. 4, p. 571–583, 2007.
- BUDGEN, D.; BURN, A. J.; KITCHENHAM, B. Reporting computing projects through structured abstracts: A quasi-experiment. **Empirical Software Engineering**, v. 16, n. 2, p. 244–277, 2011.
- BUI, D. D. A. et al. Extractive text summarization system to aid data extraction from full text in systematic review development. **Journal of Biomedical Informatics**, v. 64, p. 265–272, 2016.
- CAPUTO, G. M.; EBECKEN, N. F. F. Concept map construction applying natural language processing on text extracted from e-commerce web pages. In: **3rd World Congress on Nature and Biologically Inspired Computing - NaBIC**. Salamanca, Spain: [s.n.], 2011. p. 409–414.
- CEAUSU, V.; DESPRÉS, S. Learning term to concept mapping through verbs: A case study. In: **CEUR Workshop Proceedings**. Whistler, Canada: [s.n.], 2007. v. 289.
- CERQUEIRA, A. J. d. D. d. O. et al. **Implementação de Buscas Utilizando Linguagem Natural Através de Algoritmos Adaptativos**. São paulo: Escola Politécnica da Universidade de São Paulo, 2010. 1–140 p.
- CHAMPAIGN, J.; MCCALLA, G. Educational Question Answering Motivated by Question-Specific Concept Maps. **Lecture Notes in Computer Science**, v. 9112, n. June, p. 883, 2015.
- CHEN, N. S. et al. Mining e-Learning domain concept map from academic articles. **Computers and Education**, v. 50, p. 1009–1021, 2008.
- COFFEY, J. W. et al. Automated Concept Map Generation from Service-Oriented Architecture Artifacts. In: **5th International Conference on Concept Mapping**. Valletta, Malta: [s.n.], 2012. v. 1, p. 49–56.
- CRUZES, D. S.; DYBÅ, T. Synthesizing evidence in software engineering research. In: **10th International Symposium on Empirical Software Engineering and Measurement -ESEM**. Bolzano-Bozen, Italy: [s.n.], 2010. p. 1.
- DAYRELL, C. et al. Rhetorical move detection in english abstracts: Multi-label sentence classifiers and their annotated corpora. In: **LREC**. [S.l.: s.n.], 2012. p. 1604–1609.
- DOGAN, B.; DIKBIYIK, E. **OPCOMITS: Developing an adaptive and intelligent web based educational system based on concept map model**. 2016. 1–16 p.

- DUMAIS, S. T. et al. Using latent semantic analysis to improve access to textual information. In: ACM. **Proceedings of the SIGCHI conference on Human factors in computing systems**. [S.I.], 1988. p. 281–285.
- DYBA, T.; DINGSOYR, T. Empirical studies of agile software development: A systematic review. **Information and Software Technology**, v. 50, n. 9–10, p. 833–859, 2008.
- DYBA, T.; DINGSOYR, T.; HANSSEN, G. K. Applying systematic reviews to diverse study types: An experience report. In: **1st International Symposium on Empirical Software Engineering and Measurement (ESEM)**. Madrid, Spain: IEEE Computer Society, 2007. p. 225–234.
- ENDO, A. T.; SIMAO, A. Evaluating test suite characteristics, cost, and effectiveness of fsm-based testing methods. **Information and Software Technology**, v. 55, n. 6, p. 1045 – 1062, 2013. ISSN 0950-5849.
- FAILY, S. et al. Requirements Sensemaking Using Concept Maps. In: **International Conference on Human-Centred Software Engineering**. Toulouse, France: [s.n.], 2012. p. 217–232.
- FELIZARDO, K. R. et al. A Visual Analysis Approach to Validate the Selection Review of Primary Studies in Systematic Reviews. **Information and Software Technology**, v. 54, p. 1079–1091, 2012.
- FELIZARDO, K. R. et al. Using Forward Snowballing to update Systematic Reviews in Software Engineering. **Proceedings of the 10th ACM/IEEE International Symposium on Empirical Software Engineering and Measurement - ESEM '16**, n. September, p. 1–6, 2016.
- FELIZARDO, K. R.; TAKEMIYA, S. H.; SOUZA, E. F. Analyzing the Use of Graphical Abstracts to Support Study Selection in Secondary Studies. In: **Experimental Software Engineering - ESELAW 2017**. Buenos Aires - Argentina: [s.n.], 2017. p. 1–10.
- GRAUDINA, V.; GRUNDSPEKIS, J. Algorithm of Concept Map Transformation to Ontology for Usage in Intelligent Knowledge Assessment System. In: **11th International Conference on Computer Systems and Technologies - CompSysTech**. Vienna, Austria: [s.n.], 2011. p. 109–114.
- GURUPUR, V. P.; Pankaj Jain, G.; RUDRARAJU, R. Evaluating student learning using concept maps and Markov chains. **Expert Systems with Applications**, v. 42, n. 7, p. 3306–3314, 2015.
- GURUPUR, V. P. et al. Semantic requirements sharing approach to develop software systems using concept maps and information entropy: A Personal Health Information System example. **Advances in Engineering Software**, v. 70, p. 25–35, 2014.
- HUANG, H. S. et al. Effects of multidimensional concept maps on fourth graders' learning in web-based computer course. **Computers and Education**, v. 58, p. 863–873, 2012.
- HUBWIESER, P.; MÜHLING, A. What students (should) know about object oriented programming. In: **7th international workshop on Computing education research - ICER**. Rhode Island, USA: [s.n.], 2011. p. 77–84.
- JARVENPAA, S. L.; DICKSON, G. W. Graphics and managerial decision making: research-based guidelines. **Communications of the ACM**, v. 31, n. 6, p. 764–774, 1988.

- JUNQUEIRA, M. M.; Silva, Patrícia, A.; MAXIMIANO, F. A. A methodology to change student-written texts into representative concept maps. In: **6th International Conference in Concept Mapping**. Santos, Brazil: [s.n.], 2014. p. 7–14.
- JURAFSKY, D.; MARTIN, J. H. **Speech and language processing : an introduction to natural language processing, computational linguistics, and speech recognition**. Upper Saddle River, N.J.: Pearson Prentice Hall, 2009.
- KE, Z. Research on the Approach of Automatic Construct Concept Maps from Online Course. **Information Technology Journal**, v. 12, n. 24, p. 8020–8024, 2013.
- KITCHENHAM, B. Procedures for performing systematic reviews. **Keele, UK**, v. 33, n. TR/SE-0401, p. 28, 2004.
- KITCHENHAM, B.; CHARTERS, S. **Guidelines for performing Systematic Literature Reviews in Software Engineering**. Durham, UK, 2007. v. 2, 1051 p.
- KITCHENHAM, B. et al. Length and readability of structured software engineering abstracts. **IET Software**, v. 2, n. 1, p. 37, 2008.
- KITCHENHAM, B. A.; DYBA, T.; JORGENSEN, M. Evidence-based software engineering. In: **26th International Conference on Software Engineering - ICSE**. Scotland, UK: [s.n.], 2004. p. 273–281.
- KOF, L. et al. Concept Mapping as a Means of Requirements Tracing. In: **The Journal of nursing education**. Sidney, Australia: [s.n.], 2007. v. 46, p. 232–234.
- KOWATA, J. H. et al. Concept maps core elements candidates recognition from text. In: **4th International Conference on Concept Mapping**. Viña del Mar, Chile: [s.n.], 2010. p. 120–127.
- LIKERT, R. A technique for the measurement of attitudes. **Archives of psychology**, 1932.
- MARTIN-RODILLA, P.; GONZALEZ-PEREZ, C. Extracting Static and Dynamic Model Elements from Textual Specifications in Humanities. In: **4th International Conference on Technological Ecosystems for Enhancing Multiculturality**. Salamanca, Spain: [s.n.], 2016. p. 949–955.
- MARTINEZ-MALDONADO, R.; YACEF, K.; KAY, J. TSCL: A conceptual model to inform understanding of collaborative learning processes at interactive tabletops. **International Journal of Human-Computer Studies**, v. 83, p. 62–82, 2015.
- NAKAGAWA, E. et al. **Revisão Sistemática da Literatura em Engenharia de Software: Teoria e Prática**. Brazil: Elsevier Brasil, 2017. ISBN 9788535285970.
- NOVAK, J. D. **The theory underlying concept maps and how to construct and use them**. [S.I.], 2003. 1–6 p.
- NOVAK, J. D.; CAÑAS, A. J. The origins of the concept mapping tool and the continuing evolution of the tool. **Information Visualization**, v. 5, n. 3, p. 175–184, 2006.
- NOVAK, J. D.; CAÑAS, a. J. **The Theory Underlying Concept Maps and How to Construct and Use Them**. Pensacola, USA, 2008. 1–36 p.

- NOVAK, J. D.; MUSONDA, D. A Twelve-Year Longitudinal Study of Science Concept Learning. **American Educational Research Journal**, v. 28, n. 1, p. 117–153, jan 1991.
- NUGUMANOVA, A. et al. Automatic Generation of Concept Maps based on Collection of Teaching Materials. In: **4th International Conference on Data Management Technologies and Applications**. Colmar, France: [s.n.], 2015. p. 248–254.
- OLIVEIRA, A.; PEREIRA, F. C.; CARDOSO, A. Automatic reading and learning from text. In: **International Symposium on Artificial Intelligence - ISAI**. Fort Panhala, India: [s.n.], 2001. p. 1–12.
- OLNEY, A. M.; CADE, W. L.; WILLIAMS, C. **Generating Concept Map Exercises from Textbooks**. Portland, USA: [s.n.], 2011. 111-119 p.
- PETERSEN, K.; NAUMAN, B. A. Identifying strategies for study selection in systematic reviews and maps. In: **5th International Symposium on Empirical Software Engineering and Measurement (ESEM)**. Banff, Canada: IEEE Computer Society, 2011. p. 1–10.
- PETERSEN, K.; VAKKALANKA, S.; KUZNIARZ, L. Guidelines for conducting systematic mapping studies in software engineering: An update. **Information and Software Technology**, v. 64, p. 1–18, 2015.
- PUNURU, J.; CHEN, J. Learning non-taxonomical semantic relations from domain texts. **Journal of Intelligent Information Systems**, v. 38, n. 1, p. 191–207, 2012.
- QASIM, I. et al. Concept map construction from text documents using affinity propagation. **Journal of Information Science**, v. 39, p. 719–736, 2013.
- RICHARDSON, R.; FOX, E. A. Using Concept Maps in NDLTD as a Cross-Language Summarization Tool for Computing - Related ETDs Monolingual ETD Experiment. In: **5th ACM/IEEE Joint Conference on Digital Libraries JCDL**. Denver, USA: [s.n.], 2005. p. 711.
- RIGBY, S. et al. The efficacy of cross-discipline representations for ill-defined concepts. In: **10th ACM conference on SIG-information technology education - SIGITE**. Fairfax, USA: [s.n.], 2009. p. 229–234.
- SANTOS, V. et al. Analyzing the use of concept maps in computer science: a systematic mapping study. **Informatics and Education**, v. 16, n. 2, p. 257–288, 2017.
- SIEN, V. Integrating concept mapping techniques into object-oriented analysis and design curriculum. **international Journal of Learning**, v. 17, n. 2, p. 109–124, 2010.
- SILVA, C. P. d.; MORAIS, J. M.; MONTEIRO, D. C. Text2MARK: A text mining tool in the aid of knowledge representation. In: **International Conference on Intelligent Systems Design and Applications, ISDA**. Okinawa, Japan: [s.n.], 2014. p. 236–241.
- SNIDER, D. et al. Using concept maps to introduce software security assurance cases. **Cross-Talk**, v. 5, p. 211–224, jan 2014.
- TAKEMIYA, S. hesae; FELIZARDO, K. R. Dissertação de mestrado em andamento, **Uso de resumos gráficos baseados em Mapas Conceituais para apoiar a seleção de estudos primários no processo de revisão sistemática na engenharia de software**. Cornélio Procópio, Brasil: [s.n.], 2016. 1–37 p.

- TOMOTO, T. et al. Collaborative knowledge construction using concept maps for cross-cultural communication. In: **5th International Conference on Mobile Ubiquitous Computing, Systems, Services and Technologies (PECES) 3rd International Workshop on Pervasive Computing in Embedded Systems**. Lisboa, Portugal: [s.n.], 2011. p. 180–186.
- VAISHNAVI, P. C.; PRASATH, A. R.; SHOBANA, B. T. Concept Map Based Approach for analysing Student's Level of Learning. In: **National Conference on Communication and Informatics**. Guwahati, India: [s.n.], 2016. p. 70–78.
- VALERIO, A.; LEAKE, D. B.; CAÑAS, A. J. Using Automatically Generated Concept Maps for Document Understanding: a Human Subjects Experiment. In: **5th International Conference on Concept Mapping**. Valletta, Malta: [s.n.], 2012. v. 2, n. 1, p. 438–445.
- VILLALON, J.; CALVO, R. A. Concept maps as cognitive visualizations of writing assignments. **Educational Technology and Society**, v. 14, n. 3, p. 16–27, 2011.
- WANG, W. et al. Self-associated concept mapping for representation, elicitation and inference of knowledge. **Knowledge-Based Systems**, v. 21, p. 52–61, 2008.
- WANG, W. M. et al. Mining knowledge from natural language texts using fuzzy associated concept mapping. **Information Processing and Management**, v. 44, n. 5, p. 1707–1719, 2008.
- WANG, W. M. et al. A computational narrative construction method with applications in organizational learning of social service organizations. **Expert Systems with Applications**, Elsevier Ltd, v. 36, n. 4, p. 8093–8102, 2009.
- WILLIAMS, C. B. et al. Advancing Personalized Learning Via an Adaptive concept map. In: **American Society for Engineering Education**. San Antonio, Texas: [s.n.], 2012. p. 12.
- YANG, Y. F. Automatic scaffolding and measurement of concept mapping for EFL students to write summaries. **Educational Technology and Society**, v. 18, n. 4, p. 273–286, 2015.
- YOON, W. C.; LEE, S.; LEE, S. Burst Analysis of Text Document for Automatic Concept Map Creation. In: **27th International Conference on Industrial Engineering and Other Applications of Applied Intelligent Systems - IEA/AIE**. Kaohsiung, Taiwan: [s.n.], 2014. p. 407–416.
- ZHANG, H.; MUHAMMAD, A. Systematic reviews in software engineering: An empirical investigation. **Information and Software Technology**, p. In Press, 2012.
- ZHANG, H.; MUHAMMAD, A. B. An empirical investigation of systematic reviews in software engineering. In: **5th International Symposium on Empirical Software Engineering and Measurement (ESEM)**. Banff, Canada: IEEE Computer Society, 2011. p. 1–10.
- ZHANG, H.; MUHAMMAD A. B. TELL, P. Identifying relevant studies in software engineering. **Information and Software Technology**, v. 53, n. 1, p. 625–637, 2011.
- ZOUAQ, A.; NKAMBOU, R. Building domain ontologies from text for educational purposes. **IEEE Transactions on Learning Technologies**, v. 1, n. 1, p. 49–62, 2008.
- ZOUAQ, A. et al. Document Semantic Annotation for Intelligent Tutoring Systems : a Concept Mapping Approach. In: **FLAIRS Conference**. Florida, United States: [s.n.], 2007. p. 380–385.

ZVACEK, S. M.; RESTIVO, M. T.; CHOUZAL, M. F. Visualizing understanding with concept maps. In: **5th International Conference on Interactive Collaborative Learning - ICL**. Villach, Austria: [s.n.], 2012. p. 1–5.

APÊNDICE A – TREINAMENTO

Experiment: Evaluation of the use of concept map to support the selection of primary studies in the systematic review process — TRAINING

Instructions: Please read and perform the following task.

Task 1. Evaluate the concept maps in each item considering the affirmation. You must tick in the item that represents your opinion about each affirmation.

Input: Likert scale evaluation of Concept Maps generated (Appendix A), Abstracts and Concept maps (Appendix B).

Output waited: All Concept Maps must be evaluated in Likert scale.

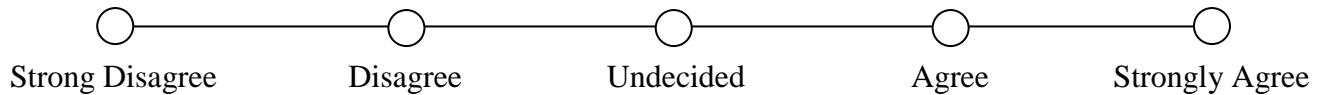
Important: Before starting make sure that you have understood how to evaluate the presented maps

Important: After you have finished the above task, answer the question on Appendix C

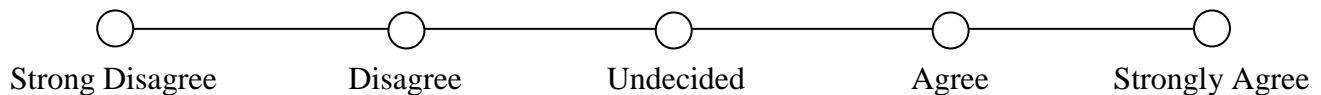
Appendix A – Questionnaire

Evaluation

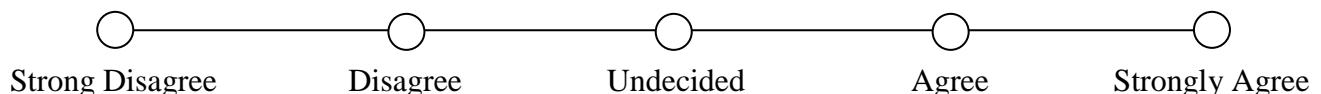
- **RQ1** – The concepts presented in the Concept Map represent ideas and information related to the element “Result” of the correspondent abstract.



- **RQ2** – The relationships (links) adequately connect the concepts related to the element “Result” of the correspondent abstract.



- **RQ3** – The Concept Map cover all information related to the element “Result” of the correspondent abstract.



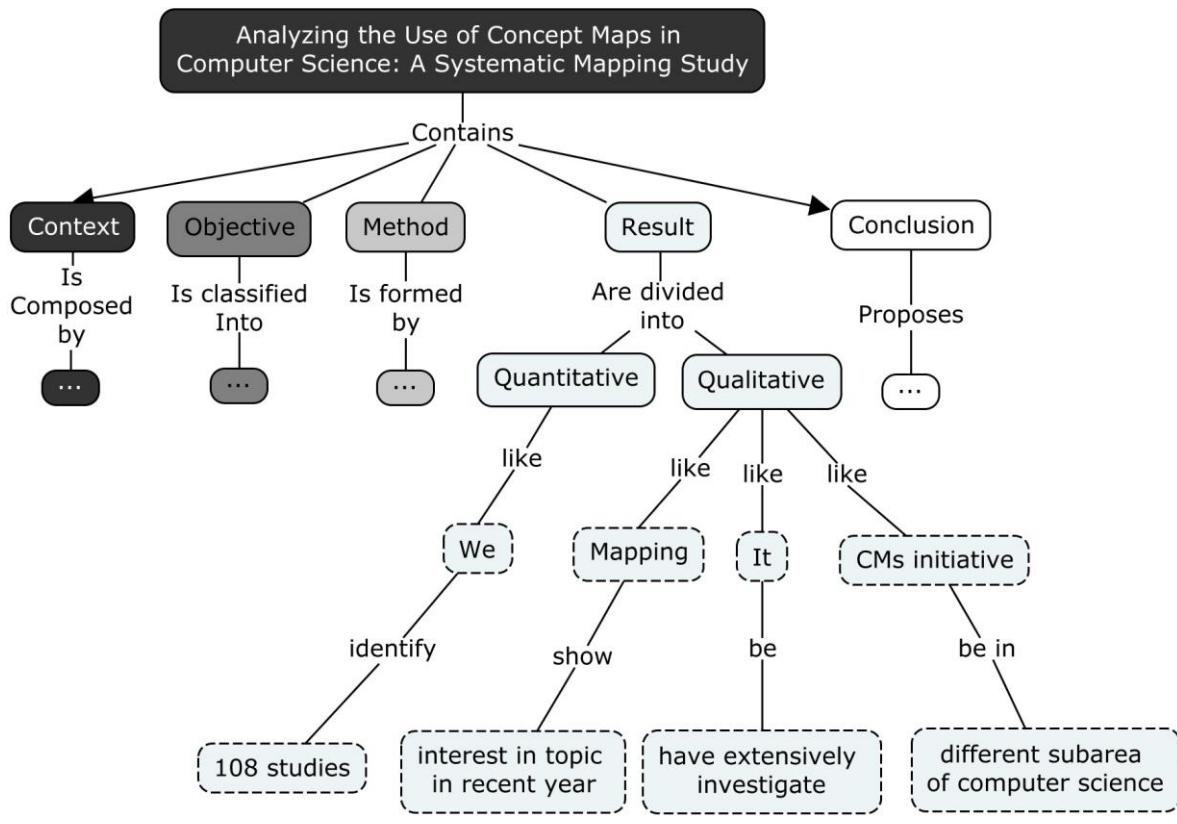
Appendix B – List of papers and their respective concept maps

Paper 1 - Analyzing the Use of Concept Maps in Computer Science: A Systematic Mapping Study

- **Abstract**

Context: concept Maps (CMs) enable the creation of a schematic representation of a domain knowledge. For this reason, CMs have been applied in different research areas, including Computer Science. Objective: the objective of this paper is to present the results of a systematic mapping study conducted to collect and evaluate existing research on CMs initiatives in Computer Science. Method: the mapping study was performed by searching five electronic databases. We also performed backward snowballing and manual search to find publications of researchers and research groups that accomplished these studies. **Results:** from the mapping study, we identified **108 studies addressing CMs initiatives in different subareas of Computer Science that were reviewed to extract relevant information to answer a set of research questions. The mapping shows an increasing interest in the topic in recent years and it has been extensively investigated due to support in teaching and learning.** Conclusions: based on our results we conclude that the use of CMs as an educational tool has been widely accepted in Computer Science.

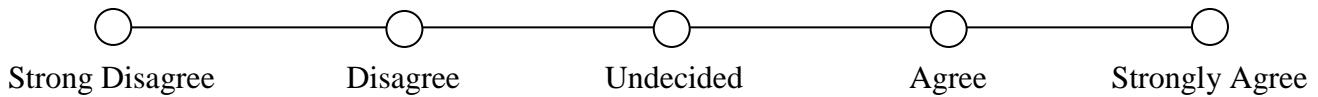
- **Concept Map**



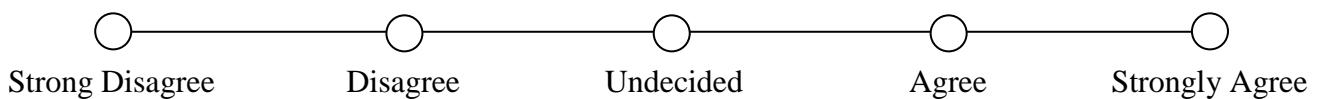
Legend: → Links → Cross-link [] Variable Concepts [] Fixed Concepts

Classification:

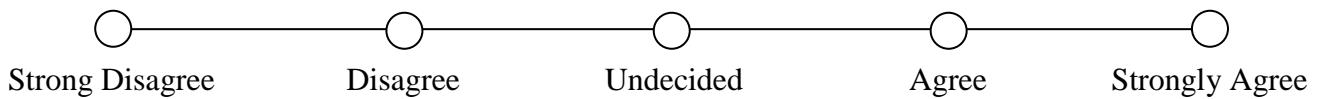
- **RQ1** – The concepts presented in the Concept Map represent ideas and information related to the element “Result” of the correspondent abstract.



- **RQ2** – The relationships (links) adequately connect the concepts related to the element “Result” of the correspondent abstract.



- **RQ3** – The Concept Map cover all information related to the element “Result” of the correspondent abstract.



End of task

APÊNDICE B – EXECUÇÃO - MCS GERADOS AUTOMATICAMENTE

Experiment: Evaluation of the use of concept map to support the selection of primary studies in the systematic review process — EXECUTION

Instructions: Please read and perform the following task.

Task 1. Evaluate the concept maps in each item considering the affirmation. You must tick in the item that represents your opinion about each affirmation.

Input: Likert scale evaluation of Concept Maps generated (Appendix A), Abstracts and Concept maps (Appendix B).

Output waited: All Concept Maps must be evaluated in Likert scale.

Important: Before starting make sure that you have understood how to evaluate the Concept Maps

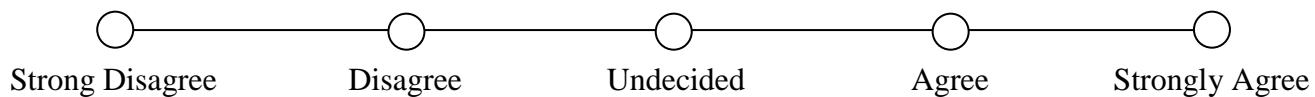
Important: After you have finished the above task, answer the question on Appendix C

Code: 1401TO

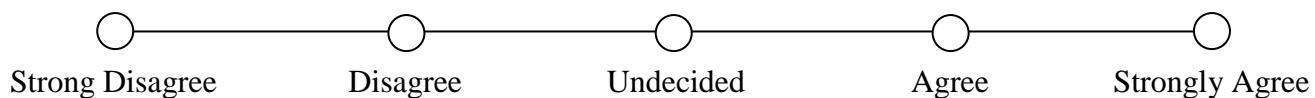
Appendix A – Questionnaire

Evaluation

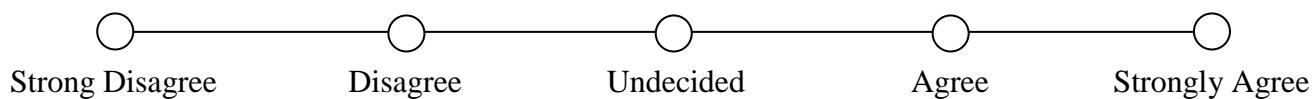
- **RQ1** – The concepts presented in the Concept Map represent ideas and information related to the element “Result” of the correspondent abstract.



- **RQ2** – The relationships (links) adequately connect the concepts related to the element “Result” of the correspondent abstract.



- **RQ3** – The Concept Map cover all information related to the element “Result” of the correspondent abstract.



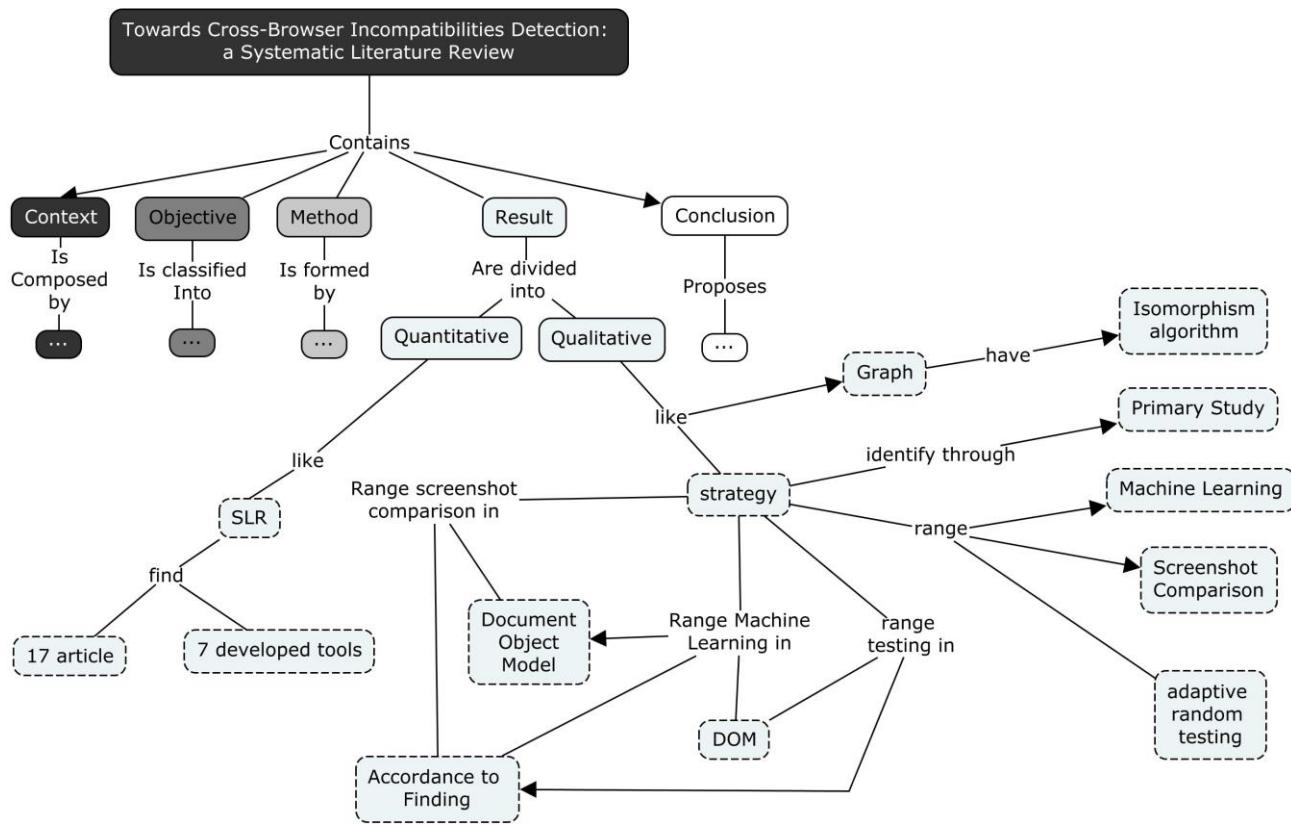
Appendix B – List of papers and their respective concept maps

Paper 1 - Towards Cross-Browser Incompatibilities Detection: a Systematic Literature Review

- **Abstract**

Cross Browser Incompatibilities (XBI) stands for compatibility issues which can be observed while rendering the same web application in different browsers. Users can interact with the Web through distinct web browsers implementations, such as: Internet Explorer, Microsoft Edge, Mozilla Firefox, Opera, Google Chrome, among others. However, with the increasing number of browser implementations and the continually evolving characteristic of web technologies lead to differences in how browsers behave and render web applications. In order to overcome this issue during the software development process, web developers must detect and fix XBIs before deploying web applications, regardless of the effort and cost required to conduct these inspections. Many developers rely on manual inspection of every web page of their applications rendered in various configuration environments (considering multiple OS platforms, browser implementations and versions) to detect XBIs, independently of the effort and cost required to conduct these inspections. This paper presents a Systematic Literature Review (SLR) on XBI automatic detection strategies published as primary studies. The SLR process was conducted with the goal of identifying distinct techniques which have been used to identify XBI, present the evolution of the developed tools, and guide future research on the topic. **In accordance to our findings, the strategies identified through the primary studies range from DOM (Document Object Model) Structure analysis, screenshot comparison, graph's isomorphism algorithms, machine learning, adaptive random testing, relative layout comparison, and static analysis. The SLR found 17 articles and 7 developed tools.**

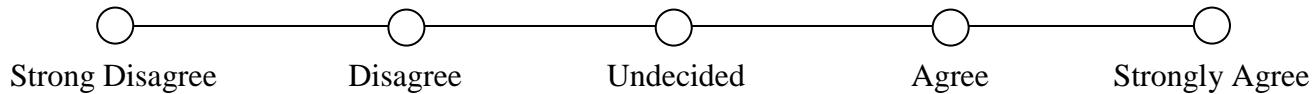
- **Concept Map**



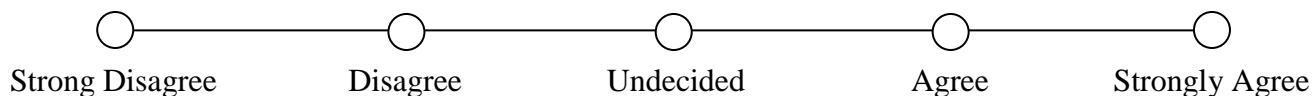
Legend: → Links → Cross-link [] Variable Concepts [] Fixed Concepts

Classification:

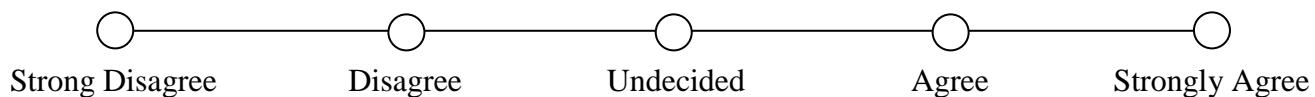
- **RQ1** – The concepts presented in the Concept Map represent ideas and information related to the element “Result” of the correspondent abstract.



- **RQ2** – The relationships (links) adequately connect the concepts related to the element “Result” of the correspondent abstract.



- **RQ3** – The Concept Map cover all information related to the element “Result” of the correspondent abstract.

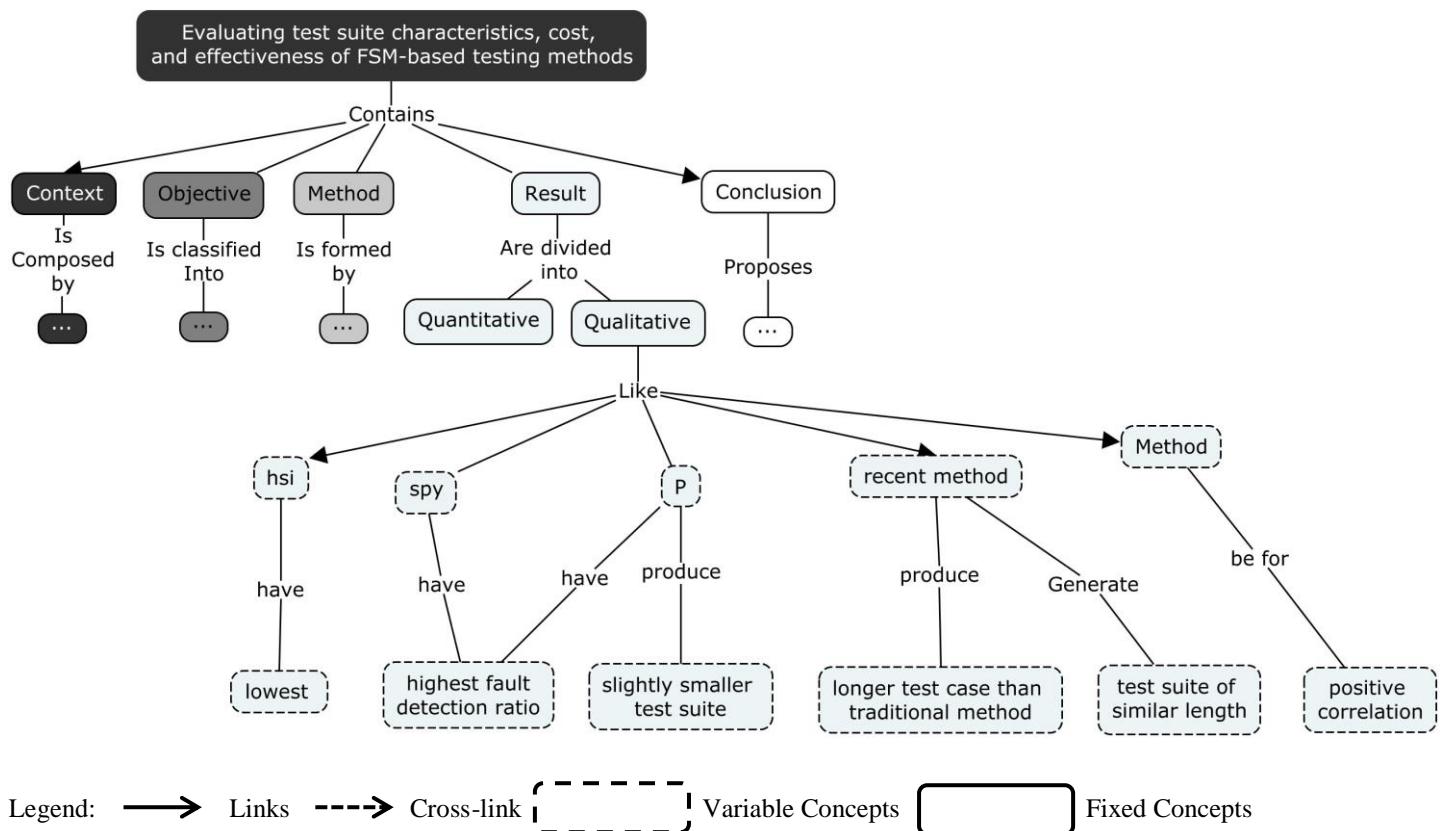


Paper 2 - Evaluating test suite characteristics, cost, and effectiveness of FSM-based testing methods

- Abstract

Testing from finite state machines has been investigated due to its well-founded and sound theory as well as its practical application. There has been a recurrent interest in developing methods capable of generating test suites that detect all faults in a given fault domain. However, the proposal of new methods motivates the comparison with traditional methods. We compare the methods that generate complete test suites from finite states machines. The test suites produced by the W, HSI, H, SPY, and P methods are analyzed in different configurations. Complete and partial machines were randomly generated varying numbers of states, inputs, outputs, and transitions. These different configurations were used to compare test suite characteristics (number of resets, test case length) and the test suite length (i.e., the sum of the length of its test cases). The fault detection ratio was evaluated using mutation testing to produce faulty implementations with an extra state. **On average, the recent methods (H, SPY, and P) produced longer test cases but smaller test suites than the traditional methods (W, HSI).** The recent methods generated test suites of similar length, though P produced slightly smaller test suites. The SPY and P methods had the highest fault detection ratios and HSI had the lowest. For all methods, there was a positive correlation between the number of resets and the test suite length and between the test case length and the fault detection ratio. The recent methods rely on fewer and longer test cases to reduce the overall test suite length, while the traditional methods produce more and shorter test cases. Longer test cases are correlated to fault detection ratio which favored SPY, though all methods have a ratio of over 92%.

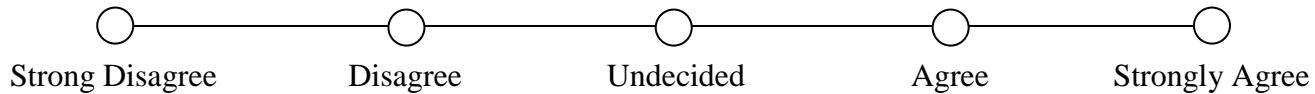
- Concept Map



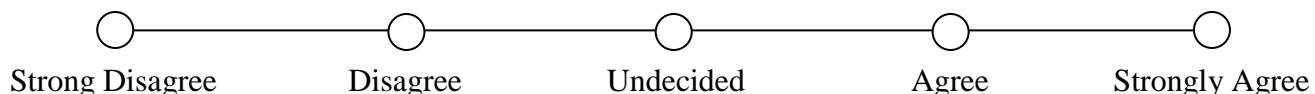
Legend: → Links → Cross-link [] Variable Concepts [] Fixed Concepts

Classification:

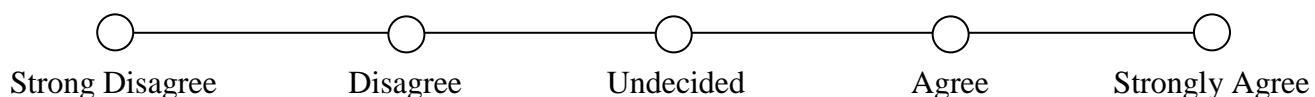
- **RQ1** – The concepts presented in the Concept Map represent ideas and information related to the element “Result” of the correspondent abstract.



- **RQ2** – The relationships (links) adequately connect the concepts related to the element “Result” of the correspondent abstract.



- **RQ3** – The Concept Map cover all information related to the element “Result” of the correspondent abstract.

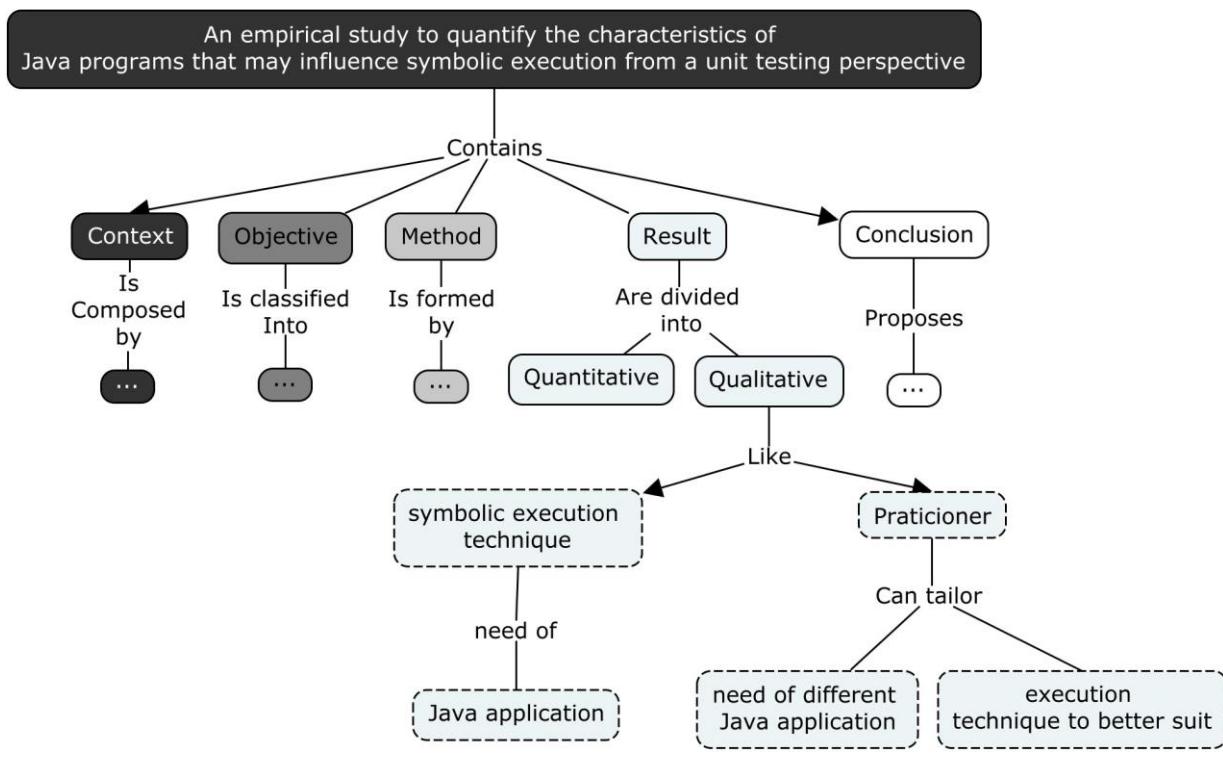


Paper 3 - An empirical study to quantify the characteristics of Java programs that may influence symbolic execution from a unit testing perspective

- **Abstract**

In software testing, a program is executed in hopes of revealing faults. Over the years, specific testing criteria have been proposed to help testers to devise test cases that cover the most relevant faulty scenarios. Symbolic execution has been used as an effective way of automatically generating test data that meet those criteria. Although this technique has been used for over three decades, several challenges remain and there is a lack of research on how often they appear in real-world applications. In this paper, we analyzed two samples of open source Java projects in order to understand the characteristics that may hinder the generation of unit test data using symbolic execution. The first sample, named SF100, is a third party corpus of classes obtained from 100 projects hosted by SourceForge. The second sample, called R47, is a set of 47 well-known and mature projects we selected from different repositories. Both samples are compared with respect to four dimensions that influence symbolic execution: path explosion, constraint complexity, dependency, and exception-dependent paths. **The results provide valuable insight into how researchers and practitioners can tailor symbolic execution techniques and tools to better suit the needs of different Java applications.**

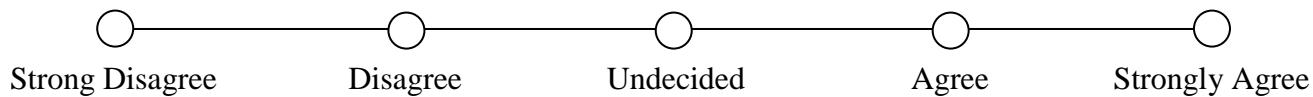
Concept Map



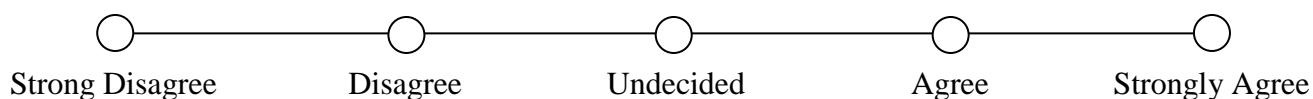
Legend: → Links → Cross-link [---] Variable Concepts Fixed Concepts

Classification:

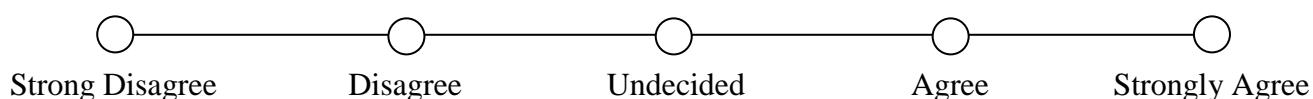
- **RQ1** – The concepts presented in the Concept Map represent ideas and information related to the element “Result” of the correspondent abstract.



- **RQ2** – The relationships (links) adequately connect the concepts related to the element “Result” of the correspondent abstract.



- **RQ3** – The Concept Map cover all information related to the element “Result” of the correspondent abstract.

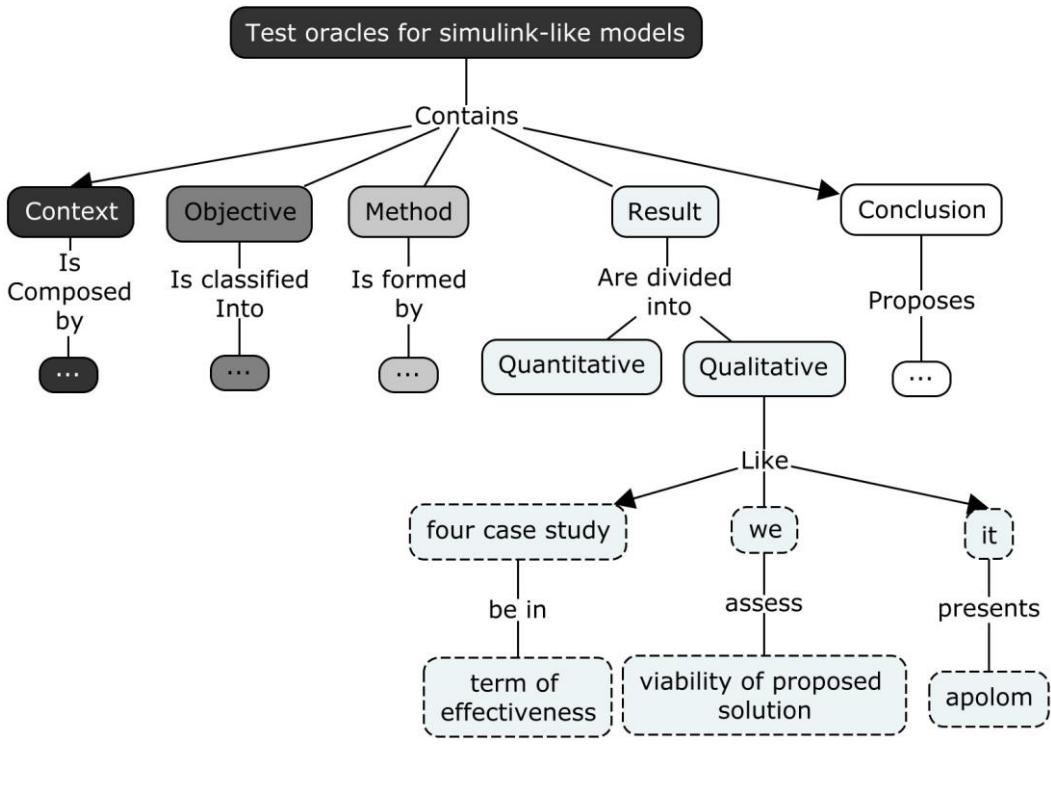


Paper 4 - Test oracles for simulink-like models

- **Abstract**

The design of embedded systems is often supported by the definition of executable models for tools like Matlab/Simulink or Scilab/Xcos. These models play a pivotal role in the development process and their correctness is thus extremely important. Many different solutions exist for the definition of suitable tests to “exercise” these models, but only a few (partial) solutions exist for assessing the quality of execution (simulation) results, that is, for defining suitable oracles. This paper addresses the problem and proposes a formal language for specifying the oracles and relating them to existing models. **It also presents Apolom, a prototype tool for checking simulation results against stated oracles.** The empirical assessment we conducted to assess the viability of the proposed solution is organized around four case studies and witnesses interesting results in terms of effectiveness, efficiency, and required resources.

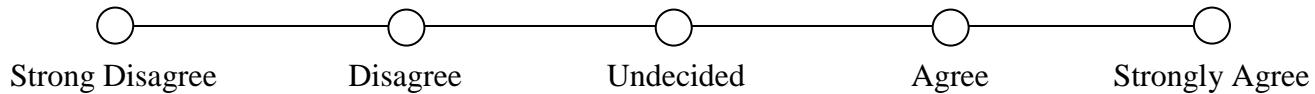
- **Concept Map**



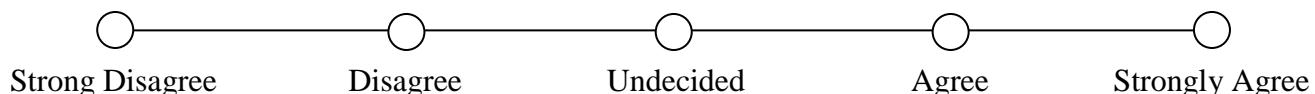
Legend: → Links ----> Cross-link [] Variable Concepts [] Fixed Concepts

Classification:

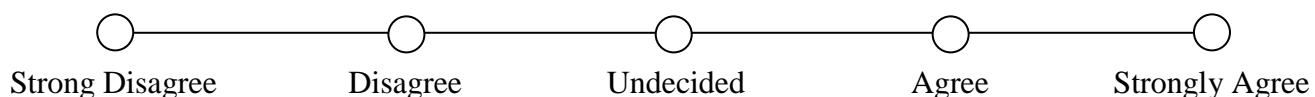
- **RQ1** – The concepts presented in the Concept Map represent ideas and information related to the element “Result” of the correspondent abstract.



- **RQ2** – The relationships (links) adequately connect the concepts related to the element “Result” of the correspondent abstract.



- **RQ3** – The Concept Map cover all information related to the element “Result” of the correspondent abstract.

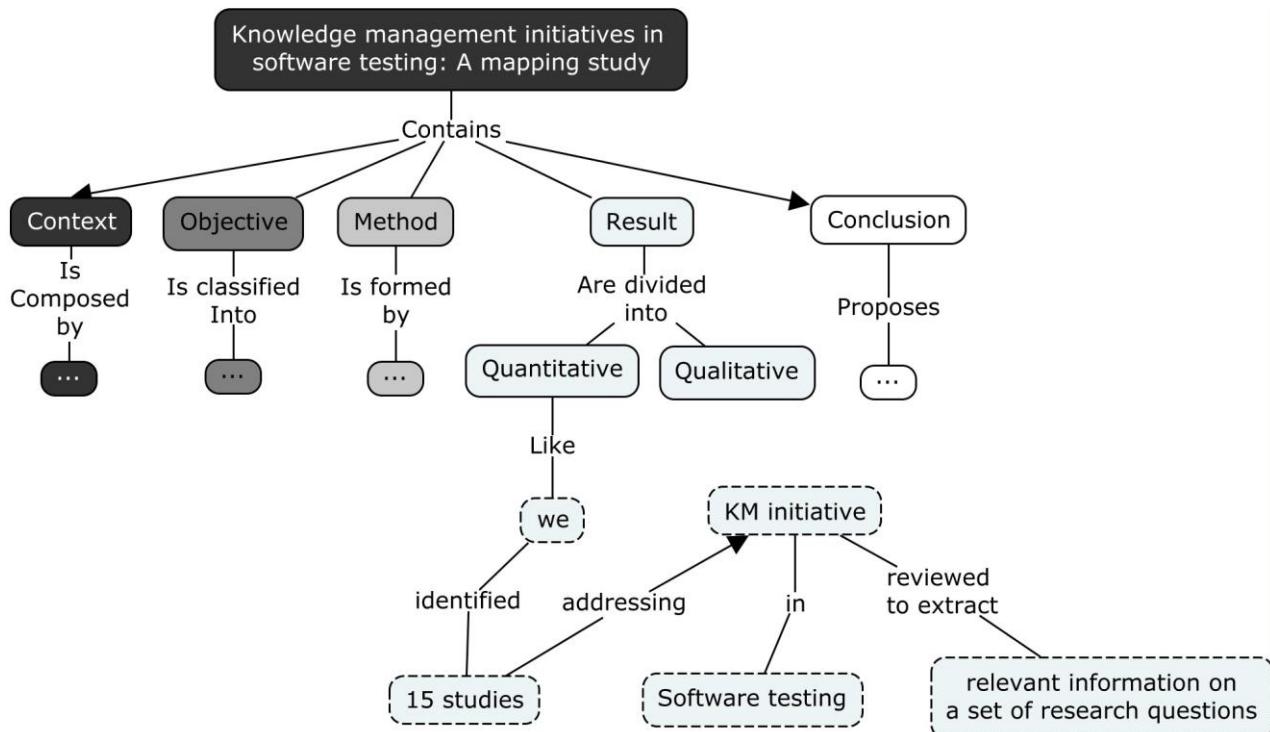


Paper 5 – Knowledge management initiatives in software testing: A mapping study

- **Abstract**

Context: Software testing is a knowledge intensive process, and, thus, Knowledge Management (KM) principles and techniques should be applied to manage software testing knowledge. **Objective:** This study conducts a survey on existing research on KM initiatives in software testing, in order to identify the state of the art in the area as well as the future research. Aspects such as purposes, types of knowledge, technologies and research type are investigated. **Method:** The mapping study was performed by searching seven electronic databases. We considered studies published until December 2013. The initial resulting set was comprised of 562 studies. From this set, a total of 13 studies were selected. For these 13, we performed snowballing and direct search to publications of researchers and research groups that accomplished these studies. **Results:** From the mapping study, we identified 15 studies addressing KM initiatives in software testing that have been reviewed in order to extract relevant information on a set of research questions. **Conclusions:** Although only a few studies were found that addressed KM initiatives in software testing, the mapping shows an increasing interest in the topic in the recent years. Reuse of test cases is the perspective that has received more attention. From the KM point of view, most of the studies discuss aspects related to providing automated support for managing testing knowledge by means of a KM system. Moreover, as a main conclusion, the results show that KM is pointed out as an important strategy for increasing test effectiveness, as well as for improving the selection and application of suited techniques, methods and test cases. On the other hand, inadequacy of existing KM systems appears as the most cited problem related to applying KM in software testing.

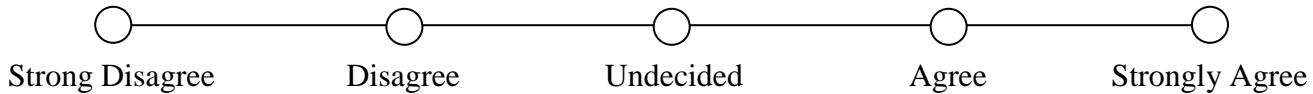
- **Concept Map**



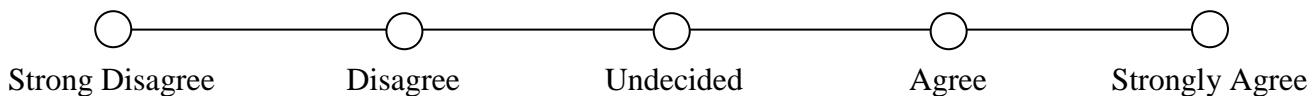
Legend: → Links → Cross-link [] Variable Concepts [] Fixed Concepts

Classification:

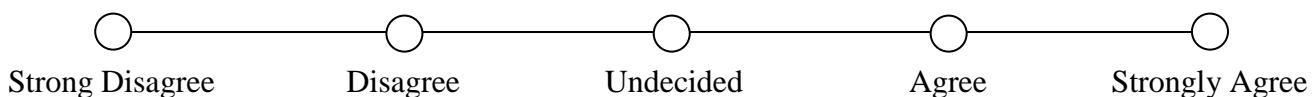
- **RQ1** – The concepts presented in the Concept Map represent ideas and information related to the element “Result” of the correspondent abstract.



- **RQ2** – The relationships (links) adequately connect the concepts related to the element “Result” of the correspondent abstract.



- **RQ3** – The Concept Map cover all information related to the element “Result” of the correspondent abstract.

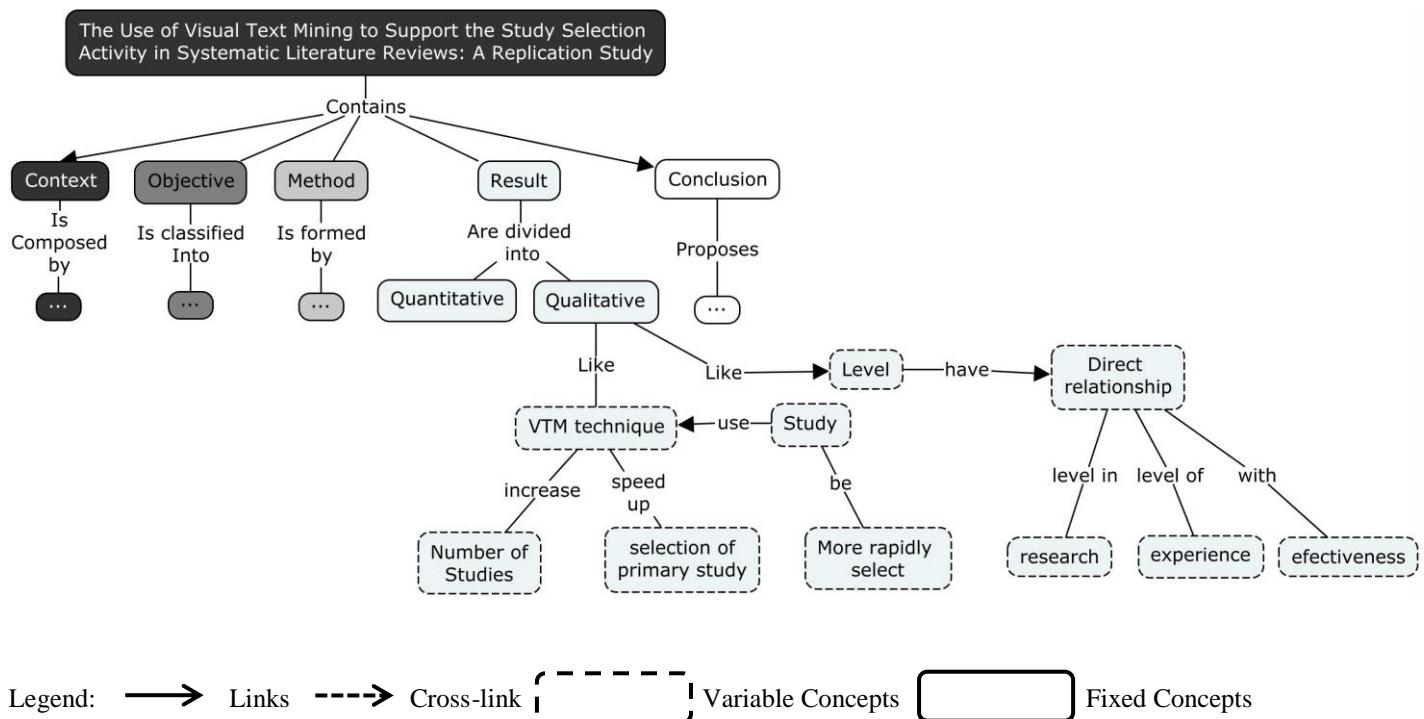


Paper 6 - The Use of Visual Text Mining to Support the Study Selection Activity in Systematic Literature Reviews: A Replication Study

- **Abstract**

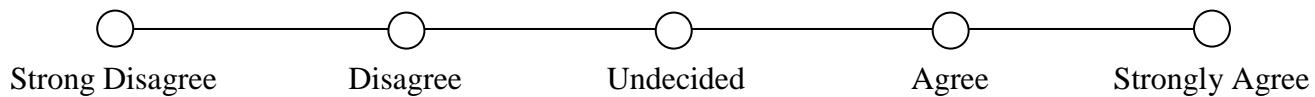
Background: Systematic literature reviews (SLRs) are an important component to identify and aggregate research evidence from different empirical studies. One of the activities associated with the SLR process is the selection of primary studies. The process used to select primary studies can be arduous, particularly when the researcher faces large volumes of primary studies. **Aim:** An experiment was conducted as a pilot test to compare the performance and effectiveness of graduate students in selecting primary studies manually and using visual text mining (VTM) techniques. This paper describes a replication study. **Method:** The same experimental design and materials of the previous experiment were used in the current experiment. **Result:** The previous experiment revealed that VTM techniques can speed up the selection of primary studies and increase the number of studies correctly included/excluded (effectiveness). The results of the replication confirmed that studies are more rapidly selected using VTM. We observed that the level of experience in researching has a direct relationship with the effectiveness. **Conclusion:** VTM techniques have proven valuable in the selection of primary studies.

- **Concept Map**

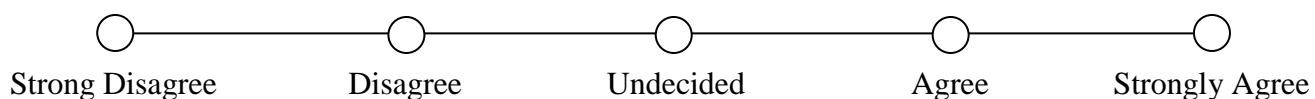


Classification:

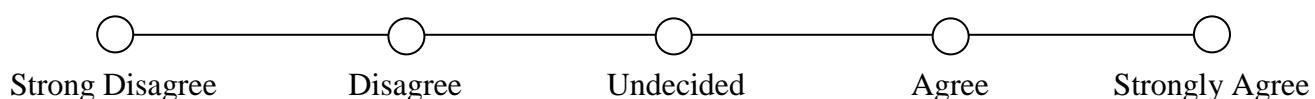
- **RQ1** – The concepts presented in the Concept Map represent ideas and information related to the element “Result” of the correspondent abstract.



- **RQ2** – The relationships (links) adequately connect the concepts related to the element “Result” of the correspondent abstract.



- **RQ3** – The Concept Map cover all information related to the element “Result” of the correspondent abstract.



End of Task.

After you have performed the above task, please answer the question on Appendix C.

Appendix C

Instructions:

Please answer the following questions.

Do you have some consideration about this research?

Thank you for your time!

APÊNDICE C – EXECUÇÃO - MCS GERADOS PELOS AUTORES

Experiment: Evaluation of the use of concept map to support the selection of primary studies in the systematic review process — EXECUTION

Instructions: Please read and perform the following task.

Task 1. Evaluate the concept maps in each affirmation.

Input: Likert scale evaluation of Concept Maps generated (Appendix A), Abstracts and Concept maps (Appendix B).

Output waited: All Concept Maps must be evaluated in Likert scale.

Important: Before starting make sure that you have understood how to evaluate the Concept Maps

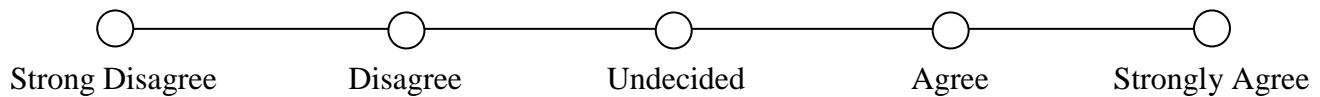
Important: After you have finished the above task, answer the question on Appendix C

Code: 1401THOR

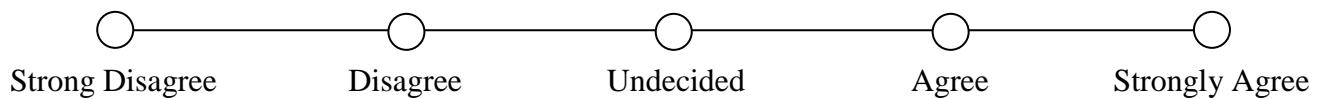
Appendix A – Questionary

Evaluation

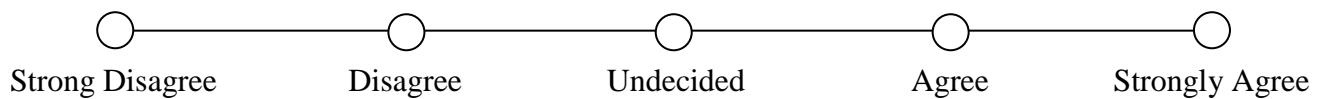
- **RQ1** – The concepts presented in the Concept Map represent ideas and information related to the element “Result” of the correspondent abstract.



- **RQ2** – The relationships (links) adequately connect the concepts related to the element “Result” of the correspondent abstract.



- **RQ3** – The Concept Map cover all information related to the element “Result” of the correspondent abstract.



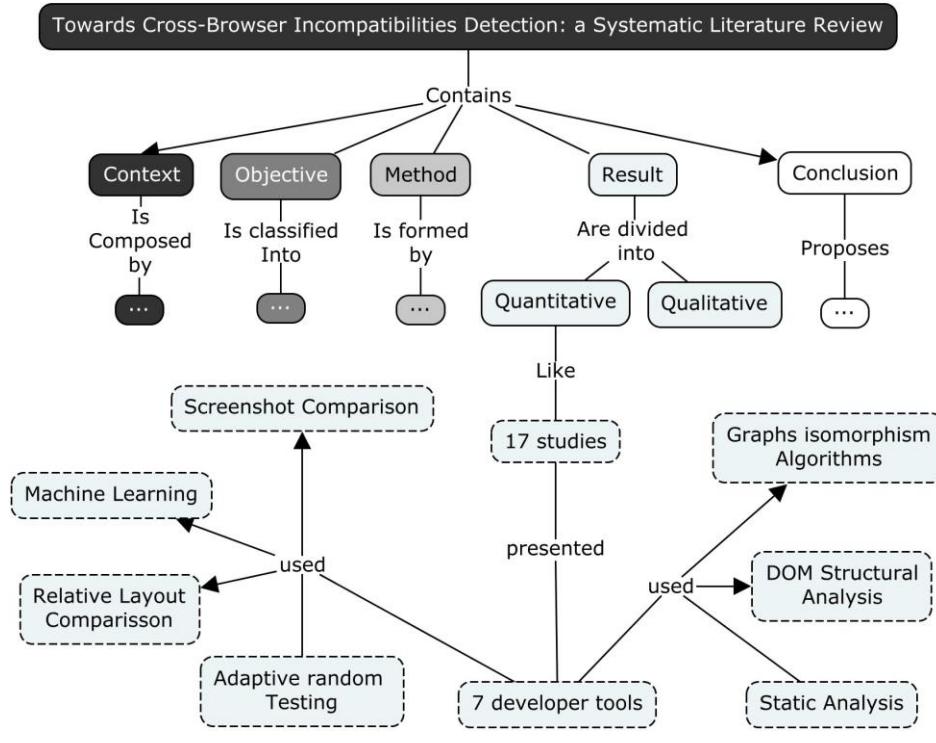
Appendix B – List of papers and their respective concept maps

Paper 1 - Towards Cross-Browser Incompatibilities Detection: a Systematic Literature Review

- **Abstract**

Cross Browser Incompatibilities (XBI) stands for compatibility issues which can be observed while rendering the same web application in different browsers. Users can interact with the Web through distinct web browsers implementations, such as: Internet Explorer, Microsoft Edge, Mozilla Firefox, Opera, Google Chrome, among others. However, with the increasing number of browser implementations and the continually evolving characteristic of web technologies lead to differences in how browsers behave and render web applications. In order to overcome this issue during the software development process, web developers must detect and fix XBIs before deploying web applications, regardless of the effort and cost required to conduct these inspections. Many developers rely on manual inspection of every web page of their applications rendered in various configuration environments (considering multiple OS platforms, browser implementations and versions) to detect XBIs, independently of the effort and cost required to conduct these inspections. This paper presents a Systematic Literature Review (SLR) on XBI automatic detection strategies published as primary studies. The SLR process was conducted with the goal of identifying distinct techniques which have been used to identify XBI, present the evolution of the developed tools, and guide future research on the topic. **In accordance to our findings, the strategies identified through the primary studies range from DOM (Document Object Model) Structure analysis, screenshot comparison, graph's isomorphism algorithms, machine learning, adaptive random testing, relative layout comparison, and static analysis. The SLR found 17 articles and 7 developed tools.**

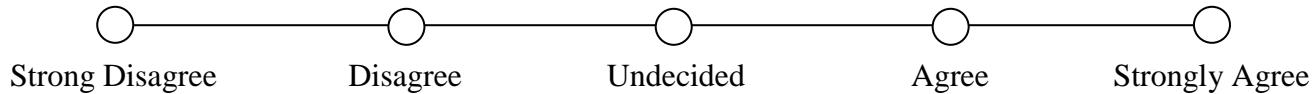
- **Concept Map**



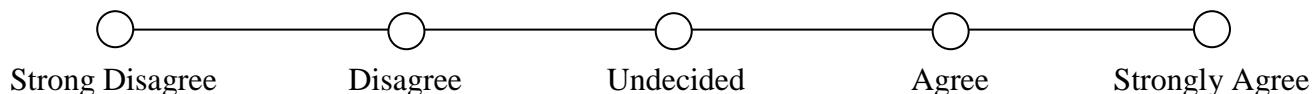
Legend: → Links → Cross-link [] Variable Concepts [] Fixed Concepts

Classification:

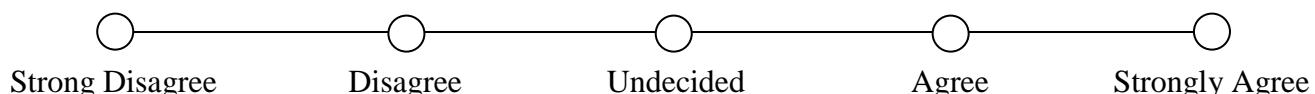
- **RQ1** – The concepts presented in the Concept Map represent ideas and information related to the element “Result” of the correspondent abstract.



- **RQ2** – The relationships (links) adequately connect the concepts related to the element “Result” of the correspondent abstract.



- **RQ3** – The Concept Map cover all information related to the element “Result” of the correspondent abstract.

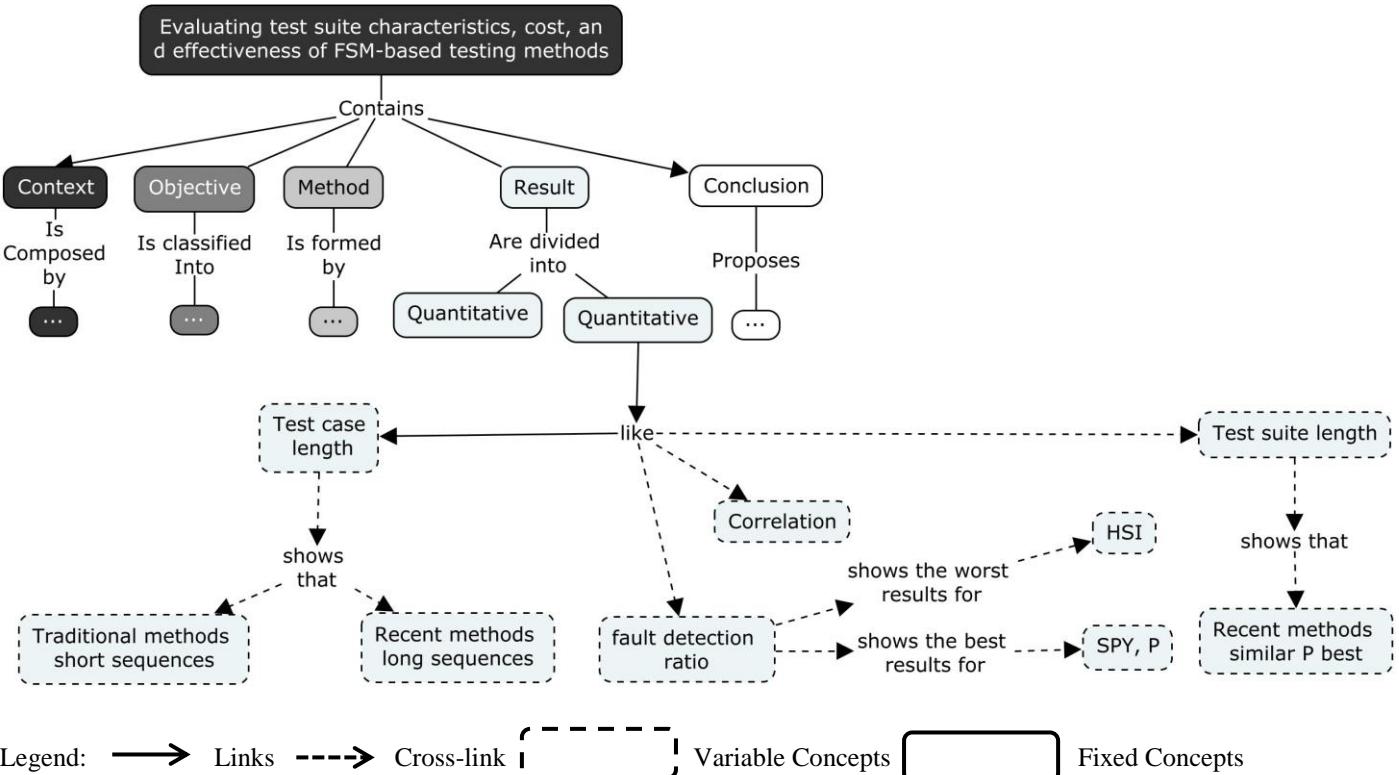


Paper 2 - Evaluating test suite characteristics, cost, and effectiveness of FSM-based testing methods

- Abstract

Testing from finite state machines has been investigated due to its well-founded and sound theory as well as its practical application. There has been a recurrent interest in developing methods capable of generating test suites that detect all faults in a given fault domain. However, the proposal of new methods motivates the comparison with traditional methods. We compare the methods that generate complete test suites from finite states machines. The test suites produced by the W, HSI, H, SPY, and P methods are analyzed in different configurations. Complete and partial machines were randomly generated varying numbers of states, inputs, outputs, and transitions. These different configurations were used to compare test suite characteristics (number of resets, test case length) and the test suite length (i.e., the sum of the length of its test cases). The fault detection ratio was evaluated using mutation testing to produce faulty implementations with an extra state. **On average, the recent methods (H, SPY, and P) produced longer test cases but smaller test suites than the traditional methods (W, HSI).** The recent methods generated test suites of similar length, though P produced slightly smaller test suites. The SPY and P methods had the highest fault detection ratios and HSI had the lowest. For all methods, there was a positive correlation between the number of resets and the test suite length and between the test case length and the fault detection ratio. The recent methods rely on fewer and longer test cases to reduce the overall test suite length, while the traditional methods produce more and shorter test cases. Longer test cases are correlated to fault detection ratio which favored SPY, though all methods have a ratio of over 92%.

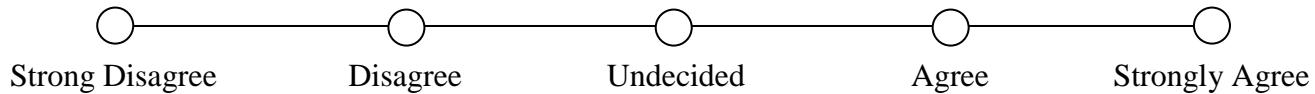
- Concept Map



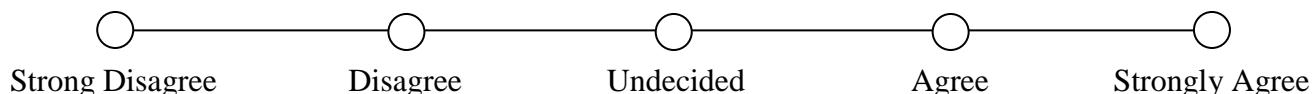
Legend: → Links → Cross-link [] Variable Concepts [] Fixed Concepts

Classification:

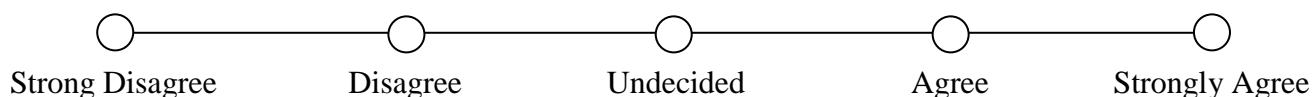
- **RQ1** – The concepts presented in the Concept Map represent ideas and information related to the element “Result” of the correspondent abstract.



- **RQ2** – The relationships (links) adequately connect the concepts related to the element “Result” of the correspondent abstract.



- **RQ3** – The Concept Map cover all information related to the element “Result” of the correspondent abstract.

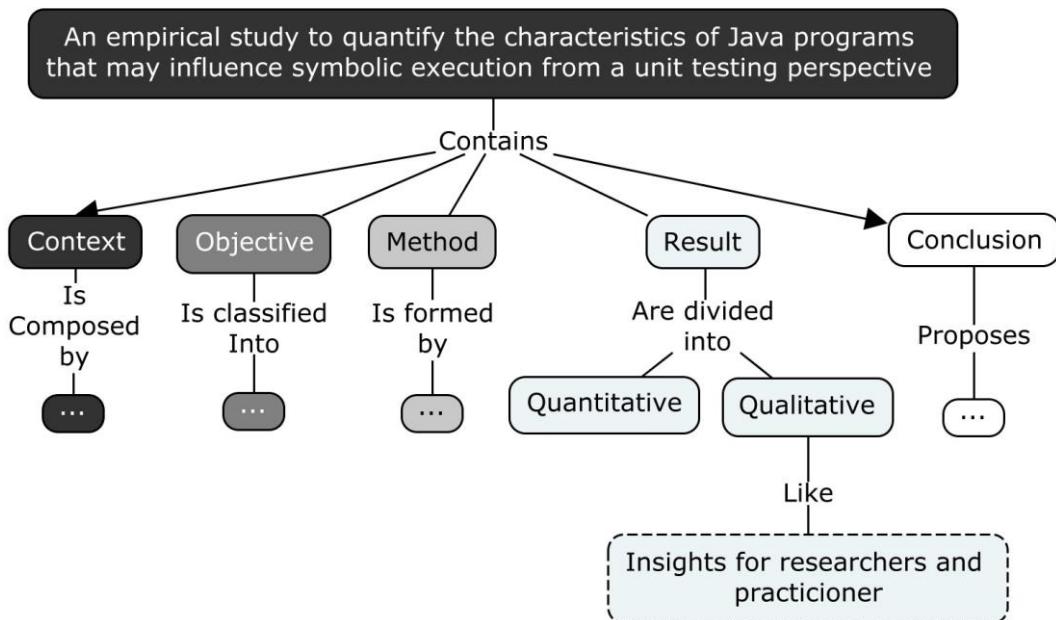


Paper 3 - An empirical study to quantify the characteristics of Java programs that may influence symbolic execution from a unit testing perspective

- **Abstract**

In software testing, a program is executed in hopes of revealing faults. Over the years, specific testing criteria have been proposed to help testers to devise test cases that cover the most relevant faulty scenarios. Symbolic execution has been used as an effective way of automatically generating test data that meet those criteria. Although this technique has been used for over three decades, several challenges remain and there is a lack of research on how often they appear in real-world applications. In this paper, we analyzed two samples of open source Java projects in order to understand the characteristics that may hinder the generation of unit test data using symbolic execution. The first sample, named SF100, is a third party corpus of classes obtained from 100 projects hosted by SourceForge. The second sample, called R47, is a set of 47 well-known and mature projects we selected from different repositories. Both samples are compared with respect to four dimensions that influence symbolic execution: path explosion, constraint complexity, dependency, and exception-dependent paths. **The results provide valuable insight into how researchers and practitioners can tailor symbolic execution techniques and tools to better suit the needs of different Java applications.**

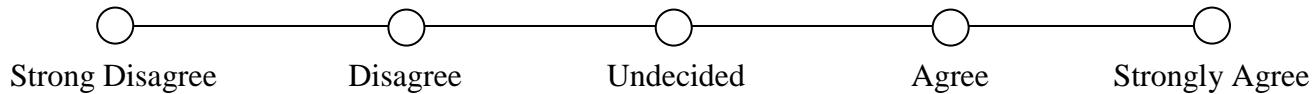
- **Concept Map**



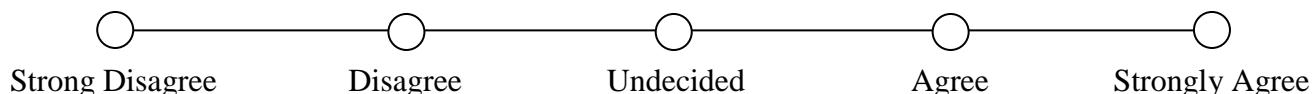
Legend: → Links → Cross-link [] Variable Concepts [] Fixed Concepts

Classification:

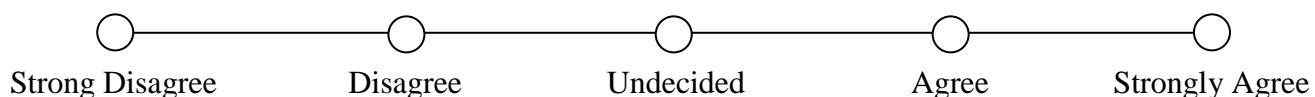
- **RQ1** – The concepts presented in the Concept Map represent ideas and information related to the element “Result” of the correspondent abstract.



- **RQ2** – The relationships (links) adequately connect the concepts related to the element “Result” of the correspondent abstract.



- **RQ3** – The Concept Map cover all information related to the element “Result” of the correspondent abstract.

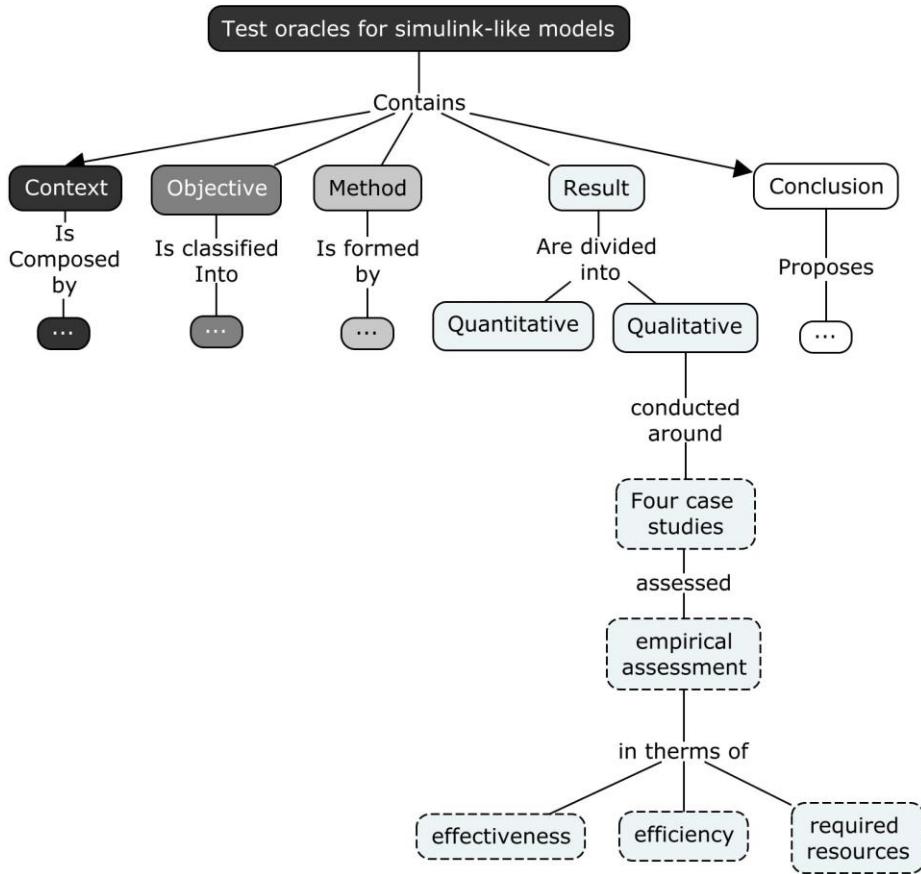


Paper 4 - Test oracles for simulink-like models

- **Abstract**

The design of embedded systems is often supported by the definition of executable models for tools like Matlab/Simulink or Scilab/Xcos. These models play a pivotal role in the development process and their correctness is thus extremely important. Many different solutions exist for the definition of suitable tests to “exercise” these models, but only a few (partial) solutions exist for assessing the quality of execution (simulation) results, that is, for defining suitable oracles. This paper addresses the problem and proposes a formal language for specifying the oracles and relating them to existing models. **It also presents Apolom, a prototype tool for checking simulation results against stated oracles.** The empirical assessment we conducted to assess the viability of the proposed solution is organized around four case studies and witnesses interesting results in terms of effectiveness, efficiency, and required resources.

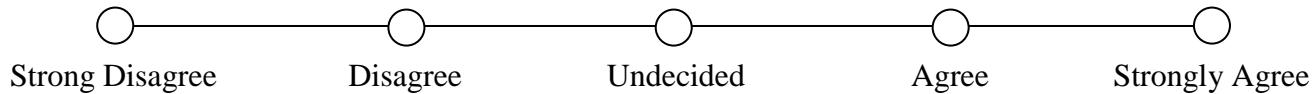
- **Concept Map**



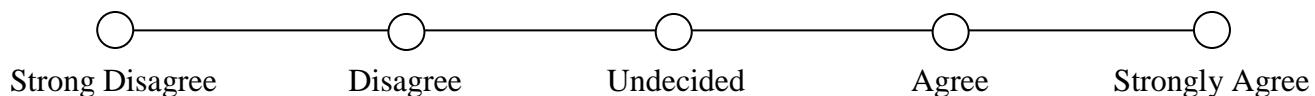
Legend: → Links ----> Cross-link [] Variable Concepts [] Fixed Concepts

Classification:

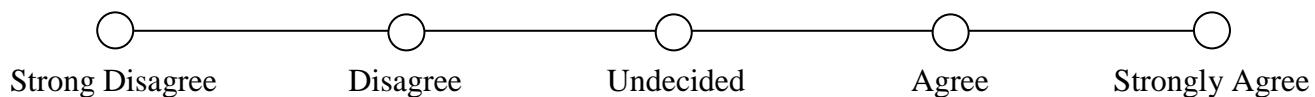
- **RQ1** – The concepts presented in the Concept Map represent ideas and information related to the element “Result” of the correspondent abstract.



- **RQ2** – The relationships (links) adequately connect the concepts related to the element “Result” of the correspondent abstract.



- **RQ3** – The Concept Map cover all information related to the element “Result” of the correspondent abstract.

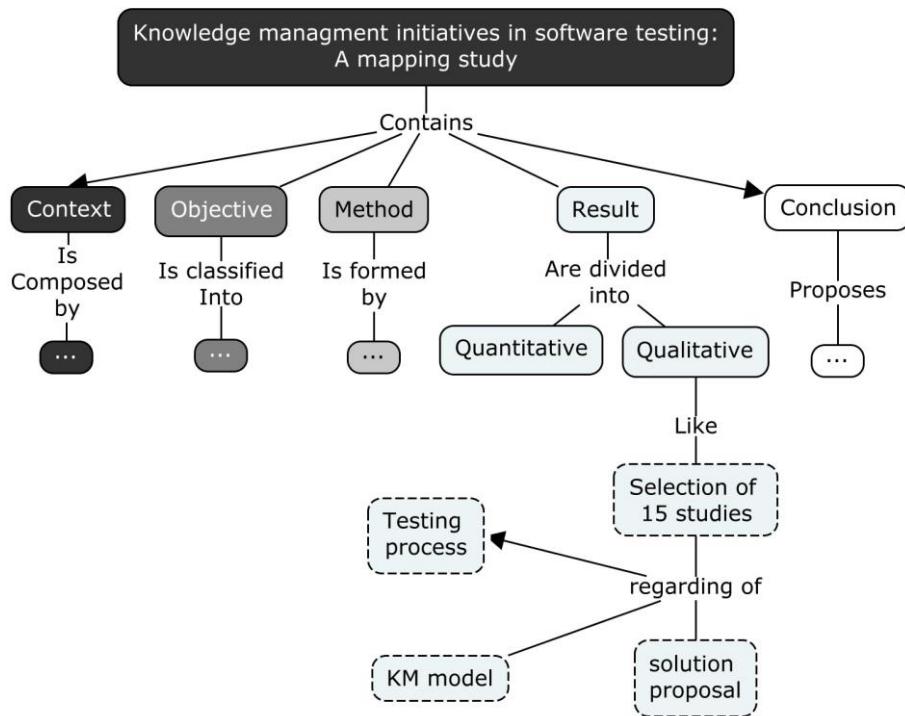


Paper 5 – Knowledge management initiatives in software testing: A mapping study

- **Abstract**

Context: Software testing is a knowledge intensive process, and, thus, Knowledge Management (KM) principles and techniques should be applied to manage software testing knowledge. **Objective:** This study conducts a survey on existing research on KM initiatives in software testing, in order to identify the state of the art in the area as well as the future research. **Aspects such as purposes, types of knowledge, technologies and research type are investigated.** **Method:** The mapping study was performed by searching seven electronic databases. We considered studies published until December 2013. The initial resulting set was comprised of 562 studies. From this set, a total of 13 studies were selected. For these 13, we performed snowballing and direct search to publications of researchers and research groups that accomplished these studies. **Results:** From the mapping study, we identified 15 studies addressing KM initiatives in software testing that have been reviewed in order to extract relevant information on a set of research questions. **Conclusions:** Although only a few studies were found that addressed KM initiatives in software testing, the mapping shows an increasing interest in the topic in the recent years. Reuse of test cases is the perspective that has received more attention. From the KM point of view, most of the studies discuss aspects related to providing automated support for managing testing knowledge by means of a KM system. Moreover, as a main conclusion, the results show that KM is pointed out as an important strategy for increasing test effectiveness, as well as for improving the selection and application of suited techniques, methods and test cases. On the other hand, inadequacy of existing KM systems appears as the most cited problem related to applying KM in software testing.

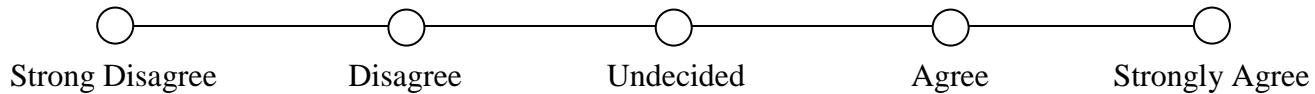
- **Concept Map**



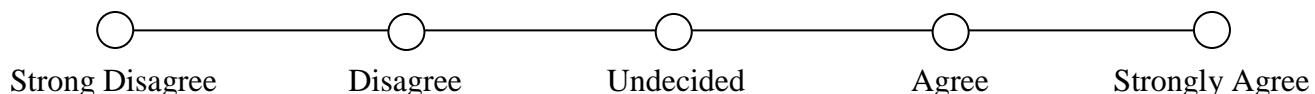
Legend: → Links ----→ Cross-link [] Variable Concepts [] Fixed Concepts

Classification:

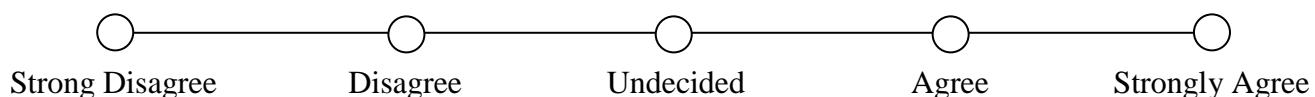
- **RQ1** – The concepts presented in the Concept Map represent ideas and information related to the element “Result” of the correspondent abstract.



- **RQ2** – The relationships (links) adequately connect the concepts related to the element “Result” of the correspondent abstract.



- **RQ3** – The Concept Map cover all information related to the element “Result” of the correspondent abstract.

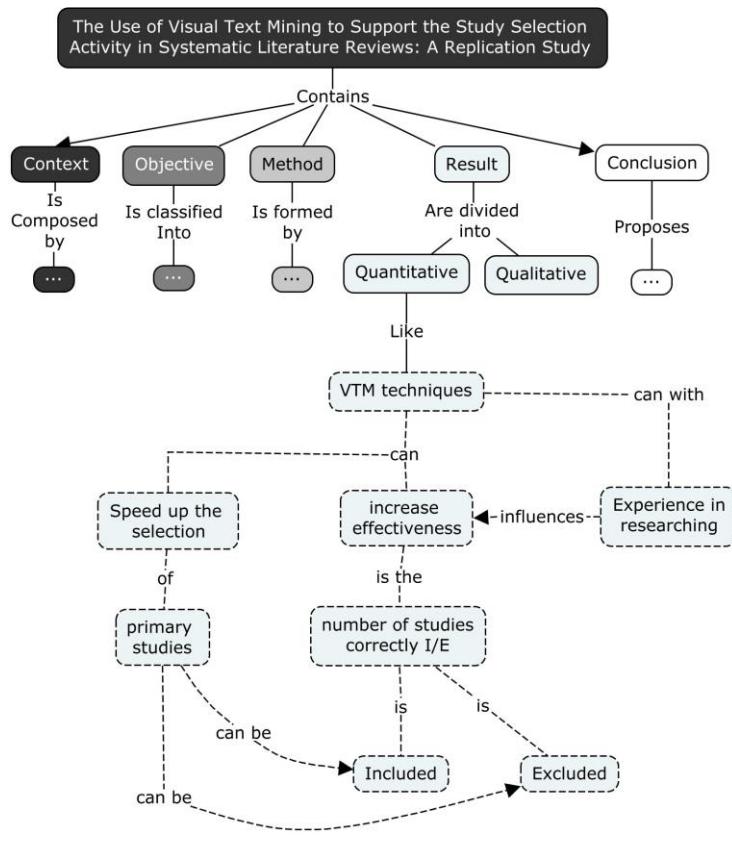


Paper 6 - The Use of Visual Text Mining to Support the Study Selection Activity in Systematic Literature Reviews: A Replication Study

- Abstract

Background: Systematic literature reviews (SLRs) are an important component to identify and aggregate research evidence from different empirical studies. One of the activities associated with the SLR process is the selection of primary studies. The process used to select primary studies can be arduous, particularly when the researcher faces large volumes of primary studies. **Aim:** An experiment was conducted as a pilot test to compare the performance and effectiveness of graduate students in selecting primary studies manually and using visual text mining (VTM) techniques. This paper describes a replication study. **Method:** The same experimental design and materials of the previous experiment were used in the current experiment. **Result:** The previous experiment revealed that VTM techniques can speed up the selection of primary studies and increase the number of studies correctly included/excluded (effectiveness). The results of the replication confirmed that studies are more rapidly selected using VTM. We observed that the level of experience in researching has a direct relationship with the effectiveness. **Conclusion:** VTM techniques have proven valuable in the selection of primary studies.

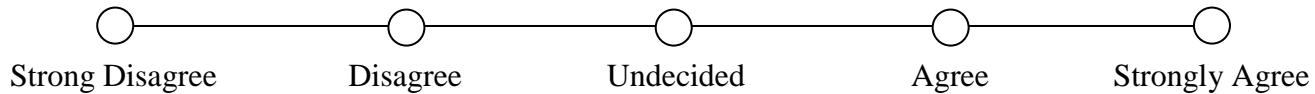
- Concept Map



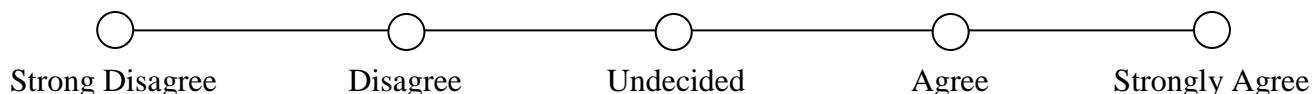
Legend: → Links - - - → Cross-link [] Variable Concepts [] Fixed Concepts

Classification:

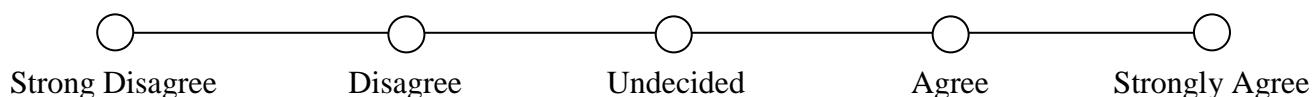
- **RQ1** – The concepts presented in the Concept Map represent ideas and information related to the element “Result” of the correspondent abstract.



- **RQ2** – The relationships (links) adequately connect the concepts related to the element “Result” of the correspondent abstract.



- **RQ3** – The Concept Map cover all information related to the element “Result” of the correspondent abstract.



End of Task.

After you have performed the above task, answer the question on Appendix C

Appendix C

Instructions:

Please answer the following questions.

Do you have some consideration about this research?

Thank you for your time!