UNIVERSIDADE ESTADUAL DE CAMPINAS

FACULDADE DE TECNOLOGIA

Leonardo Janeis de Melo

Comparação por similaridade de conhecimento representado por meio de Mapas Conceituais Estendidos

Limeira, 2021

Leonardo Janeis de Melo

Comparação por similaridade de conhecimento representado por meio de Mapas Conceituais Estendidos

Dissertação apresentada à Faculdade de Tecnologia da Universidade Estadual de Campinas como parte dos requisitos para a obtenção do título de Mestre em Tecnologia, na área de Sistemas de Informação e Comunicação.

Orientadora: Prof[a]. Dr[a]. Gisele Busichia Baioco

Este exemplar corresponde à versão final da Dissertação defendida por Leonardo Janeis de Melo e orientado pela Prof[a]. Dr[a]. Gisele Busichia Baioco.

Limeira, 2021

FICHA CATALOGRÁFICA

Ficha catalográfica
Universidade Estadual de Campinas
Biblioteca da Faculdade de Tecnologia
Felipe de Souza Bueno - CRB 8/8577

M491c

Melo, Leonardo Janeis de, 1992-
Comparação por similaridade de conhecimento representado por meio de Mapas Conceituais Estendidos / Leonardo Janeis de Melo. – Limeira, SP : [s.n.], 2021.

Orientador: Gisele Busichia Baioco.
Dissertação (mestrado) – Universidade Estadual de Campinas, Faculdade de Tecnologia.

1. Representação de conhecimento (Teoria da informação). 2. Gestão do conhecimento. 3. Algoritmos computacionais. I. Baioco, Gisele Busichia, 1970-. II. Universidade Estadual de Campinas. Faculdade de Tecnologia. III. Título.

Informações para Biblioteca Digital

Título em outro idioma: Comparison by similarity of knowledge represented by Extended Concept Maps
Palavras-chave em inglês:
Knowledge representation (Information theory)
Knowledge management
Computer algorithms
Área de concentração: Sistemas de Informação e Comunicação
Titulação: Mestre em Tecnologia
Banca examinadora:
Gisele Busichia Baioco [Orientador]
Luiz Camolesi Júnior
Maria Camila Nardini Barioni
Data de defesa: 28-05-2021
Programa de Pós-Graduação: Tecnologia

Identificação e informações acadêmicas do(a) aluno(a)
- ORCID do autor: https://orcid.org/0000-0001-6927-3742
- Currículo Lattes do autor: http://lattes.cnpq.br/0188901810020820

FOLHA DE APROVAÇÃO

Abaixo se apresentam os membros da comissão julgadora da sessão pública de defesa de dissertação para o Título de Mestre em Tecnologia na área de concentração de Sistemas de Informação e Comunicação, a que submeteu o aluno Leonardo Janeis de Melo, em 28 de maio de 2021 na Faculdade de Tecnologia - FT/ UNICAMP, em Limeira/SP.

Prof.ª. Drª. Gisele Busichia Baioco

Presidente da Comissão Julgadora

Prof. Dr. Luiz Camolesi Júnior

Faculdade de Tecnologia – FT / UNICAMP

Prof.ª. Drª. Maria Camila Nardini Barioni

Universidade Federal de Uberlândia - UFU

Ata da defesa, assinada pelos membros da Comissão Examinadora, consta no SIGA/Sistema de Fluxo de Dissertação/Tese e na Secretaria de Pós-Graduação da FT.

Agradecimentos

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

A Profa. Dra. Gisele Busichia Baioco pelos ensinamentos e orientações exatas e, principalmente, pelo incentivo e confiança para a realização deste trabalho e ao grupo de pesquisa Geicon da FT-Unicamp.

RESUMO

Mapas Conceituais Estendidos (MCE) são modelos utilizados para representação de conhecimento por meio de proposições. Um MCE é composto por uma ou mais proposições. Cada proposição é composta por dois conceitos que são posicionados em uma matriz de atributos, relacionados por um verbo e cujo relacionamento tem peso de balanceamento ou reforço. Ou seja, um MCE é uma extensão de Mapas Conceituais, agregando a estes a matriz de atributos e os pesos dos relacionamentos, de modo a proporcionar mais semântica nas representações. O MCE pode ser utilizado para diferentes aplicações, na área da gestão pública, educacional e corporativa. O objetivo do MCE é explicitar o conhecimento individual para que possa ser compartilhado. A análise visual de MCE pode exigir um grande esforço, sendo que, quanto maior a quantidade de mapas, maior será a dificuldade. Diante disso, este trabalho define um processo de comparação por similaridade de conhecimento representando por MCE como uma solução computacional que viabiliza a realização de diferentes tipos de análises. A abordagem considera o MCE como um tipo de dado complexo contendo partes estruturadas e não estruturadas, apresenta o esquema conceitual do banco de dados para seu armazenamento e um processo de comparação para diferentes possibilidades de análises de conhecimento baseadas em consultas por similaridade. O processo de comparação proposto foi implementado e aplicado a um estudo de caso de modo a verificar os resultados obtidos.

Palavras-chave: Representação de Conhecimento, Mapas Conceituais Estendidos, Mapas Conceituais, Dados Complexos, Consultas por Similaridade.

ABSTRACT

Extended Concept Maps (ECM) are models used to represent knowledge through propositions. An ECM is composed of one or more propositions. Each proposition is composed of two concepts that are placed in a matrix of attributes, related by a verb and whose relationship has a balancing or reinforcing weight. In other words, an ECM is an extension of Concept Maps, adding to them the matrix of attributes and the weights of the relationships, in order to provide more semantics in the representations. The ECM can be used for different applications, in the area of public, educational and corporate management. The purpose of ECM is to make individual knowledge explicit so that it can be shared. Visual analysis of ECM can require a great deal of effort, and the greater the amount of maps, the greater the difficulty. Therefore, this work defines a process of comparison by knowledge similarity representing by ECM as a computational solution that enables the realization of different types of analysis. The approach considers ECM as a complex data type containing structured and unstructured parts, presents the conceptual schema of the database for its storage and a comparison process for different possibilities of knowledge analysis based on similarity queries. The proposed comparison process was implemented and applied to a case study in order to verify the results obtained.

Keywords: Knowledge Representation, Extended Concept Maps, Concept Maps, Complex Data, Similarity Queries.

LISTA DE FIGURAS

LISTA DE TABELAS

SUMÁRIO

1 INTRODUÇÃO

1.1 Motivação e problemática

Organizações corporativas são ambientes de trabalho coletivo onde a construção do conhecimento coletivo é usada para adquirir vantagens competitivas sustentáveis. Para construir conhecimento coletivo nas organizações corporativas é necessário organizar e compartilhar o conhecimento individual (NONAKA; TAKEUCHI; UMEMOTO, 1996).

O conhecimento coletivo, adquirido por meio de alguma técnica de extração ou elicitação, deve ser interpretado por parte de quem for utilizá-lo. Para isso, é necessário dispor de algum tipo de representação. A representação de conhecimento pode ser realizada por um conjunto de combinações sintáticas e semânticas que torna possível explicitar o conhecimento de um indivíduo (ELMASRI; NAVATHE, 2011), para depois compartilhá-lo.

O conhecimento é encontrado em duas versões: conhecimento explícito ou conhecimento tácito. O conhecimento explícito é o conhecimento formal, estruturado e expressado de forma clara e objetiva. Pode ser formalizado através de diagramas, modelos e textos, simplificando seu compartilhamento. O conhecimento tácito é aquele que pertence ao indivíduo, sendo subjetivo e difícil de ser codificado, explicado e formalizado (NONAKA; TAKEUCHI; UMEMOTO, 1996; NONAKA, 1994).

A conversão do conhecimento tácito em explícito (codificação) é um processo benéfico para as organizações, visto que o conhecimento explícito é aquele que pode ser formalizado e compartilhado com maior facilidade. Algumas técnicas, sistemas e modelos foram criados a fim de proporcionar meios mais rápidos e efetivos de representar o conhecimento explicitado de forma estruturada. Por exemplo, a representação de conhecimento com recursos visuais, permite que atalhos sejam tomados no processo cognitivo. Esse tipo de representação pode ser feito usando conceitos da visualização da informação, explorando as capacidades humanas de percepção (CARD; MACKINLAY; SHNEIDERMAN, 1999).

A representação visual pode ser considerada mais efetiva que um texto para a comunicação de conteúdos complexos, pois o processamento mental de imagens é menos exigente cognitivamente que o processamento das regras sintáticas de um texto (VEKIRI, 2002). As principais técnicas de representação de conhecimento que usam representação visual são os Mapas Conceituais Estendidos (MCE) (ZAMBON *et al*., 2016a), Mapas Conceituais (MC) (NOVAK; CAÑAS, 2008), Redes Semânticas (LEHMANN, 1992) e Ontologias (GRUBER, 1995). Essas técnicas se assemelham por utilizarem proposições inter-relacionadas, relações semânticas e relações de causa e efeito.

Os MCE são modelos utilizados para representação de conhecimento compostos por uma ou mais proposições. Cada proposição é composta por dois conceitos. Conceito é uma representação escrita de um objeto (tangível) ou de um sentimento (intangível). Os conceitos são posicionados em uma matriz de atributos e relacionados em causa e efeito por um verbo (termo de ligação) e um peso. O peso define se a relação é de balanceamento ou reforço. O objetivo do MCE é explicitar o conhecimento individual para que sua estrutura complexa possa ser entendida.

Em um exemplo, um gestor de uma empresa pode pedir a vários funcionários que descrevam soluções individuais para um problema encontrado em um projeto, resultando em vários MCE diferentes. Considerando que todos representam soluções viáveis, a análise dos MCE deve resultar uma única solução do problema, a ser aceita por todos como a mais viável.

A análise visual de MCE pode exigir um grande esforço, visto que cada indivíduo tem seu modo singular de atribuir significados aos conceitos e estabelecer suas relações. Sendo que, quanto maior a quantidade de mapas, maior será a dificuldade. Desse modo, o uso de recursos computacionais pode auxiliar o analista.

Considerando o uso de sistemas computacionais, é possível armazenar a estrutura de MCE em bancos de dados, para posteriormente serem consultados e analisados. Entretanto, dependendo do número de mapas, a análise visual ainda exigiria esforço. Parte-se, então, para a necessidade de análise computacional, envolvendo elementos da estrutura complexa dos MCE armazenados. Desse modo, um MCE pode ser considerado um tipo de dado complexo, podendo ser consultado por seu conteúdo, a partir da comparação de suas partes componentes.

Critérios de comparação por igualdade e ordem não são úteis para dados complexos, uma vez que as estruturas de dois dados complexos dificilmente, ou quase nunca, são idênticas, a menos que seja o próprio dado. Nesses casos, a comparação pelo critério de similaridade é mais eficiente (SAMET, 2006). Sendo assim, dados complexos podem ser consultados por similaridade. Portanto, a partir do armazenamento de MCE em bancos de dados como um tipo de dado complexo, é possível realizar análises baseadas em consultas por similaridade.

De modo geral, utilizando consultas por similaridade é possível obter os MCE mais similares a um MCE de referência. Desse modo, o analista pode agrupar MCE mais similares, reduzindo a análise visual a esses agrupamentos. A base para as consultas por similaridade é realizar a comparação de dois dados complexos de um mesmo tipo a partir de suas características usando uma medida de similaridade, que pode ser obtida por meio de funções matemáticas ou mesmo algoritmos computacionais.

Vários trabalhos propuseram algoritmos e funções que realizam a comparação por similaridade de MC, Redes Semânticas e Ontologias. Porém, possuem limitações e variam seu nível de complexidade. Alguns utilizam abordagens específicas para um determinado domínio de aplicação, necessitam de especialistas para definições de parâmetros necessários ao processo de comparação ou o uso de algoritmos de processamento de linguagem natural (CHEN; LIN; CHANG, 2001; LIMONGELLI *et al.*, 2016; CALDAS; FAVERO, 2009; SILVA *et al.*, 2011). A maioria dos trabalhos, representa a estrutura de MC como grafos (DE SOUZA *et al*., 2008; LIMONGELLI *et al.*, 2016; CHEN; LIN; CHANG, 2001; IFENTHALER, 2008) para então aplicar uma abordagem de comparação, sendo alguns limitados ao uso de grafos acíclicos ou mesmo necessitam que os MC a serem comparados tenham o mesmo número de conceitos e termos de ligação (DE SOUZA *et al*., 2008; LIMONGELLI *et al.*, 2016; CHEN; LIN; CHANG, 2001). Outra restrição encontrada em algumas abordagens é a necessidade de que os rótulos dos conceitos sejam representados usando apenas uma letra do alfabeto ou uma única palavra para que a comparação seja realizada (DE SOUZA *et al*., 2008; CHEN; LIN; CHANG, 2001; LIMONGELLI *et al.*, 2016; CALDAS; FAVERO, 2009; SILVA *et al.*, 2011; IFENTHALER, 2008; SILVA *et al.*, 2011). Observa-se que para os MCE ainda não foram desenvolvidas técnicas ou métodos computacionais que comparem suas características e identifiquem semelhanças.

Este trabalho propôs a aplicação de um processo de comparação por similaridade como uma solução computacional que viabiliza a realização de diferentes tipos de análises de conhecimento representado por meio de MCE armazenados em um banco de dados.

1.2 Problema de pesquisa

Como comparar conhecimento representado por meio de MCE a fim de ser possível realizar análises computacionais?

1.3 Hipótese

Considerando que MCE são dados complexos e podem ser armazenados em banco de dados, é útil compará-los por similaridade a partir de suas características por meio da definição de uma medida de similaridade a ser utilizada em algoritmos de consultas.

1.4 Objetivo

O trabalho teve por objetivo propor um processo de comparação de MCE, utilizando como base as principais técnicas de representação de conhecimento compatíveis com a estrutura dos MCE, usando a teoria de dados complexos e consultas por similaridade.

1.5 Metodologia

A base teórica deste trabalho para tratar a abordagem de comparação de MCE, é construída utilizando princípios de revisão sistemática na literatura, buscando técnicas e algoritmos para comparação de modelos de representação gráfica de conhecimento. Foram consideradas as abordagens que atendem às necessidades de MCE para servirem como base para a proposta de solução.

As seguintes etapas compuseram a metodologia:

1. Revisão da literatura;

2. Definição de uma solução para a comparação de MCE, considerando suas características, baseando-se em técnicas e algoritmos obtidos na literatura;
3. Implementação de algoritmos de comparação por similaridade aplicáveis a MCE, utilizando a linguagem visual orientada a objetos virtual *Java*;
4. Teste da solução obtida por meio de um estudo de caso.

1.6 Organização do trabalho

Este trabalho está organizado da seguinte maneira:

- O Capítulo 2 contém a conceitualização da Representação de Conhecimento e seus modelos de representação gráfica mais similares a MCE. Posteriormente são descritas diversas técnicas de correspondência para realizar a comparação dos modelos de representação de conhecimento e algoritmos de similaridade para comparação de conceitos. O capítulo finaliza conceitualizando dados complexos e consultas por similaridade.

- O Capítulo 3 é direcionado a MCE. São descritos a parte estrutural, as características do MCE e onde já foi utilizado.

- O Capítulo 4 apresenta trabalhos que propuseram algoritmos, ferramentas e funções de similaridade, para realizar a comparação de MC em diversos contextos.

- O Capítulo 5 apresenta a abordagem para análise de MCE baseada em consultas por similaridade proposta por este trabalho.

- O Capítulo 6 apresenta o processo de comparação por similaridade de MCE, utilizando a abordagem do Capítulo 5.

- O Capítulo 7 apresenta um estudo de caso de modo a aplicar o processo de comparação por similaridade de MCE, discutindo os resultados obtidos.

- O capítulo 8 apresenta as considerações finais da pesquisa, conclusões e propostas de trabalhos futuros.

2 FUNDAMENTAÇÃO TEÓRICA

Este capítulo apresenta a fundamentação teórica de Representação de Conhecimento, conceitualizando MC e os principais modelos de representação gráfica que mais se assemelham a estrutura do MCE. Em seguida, a Seção 2.2 trata da comparação de conhecimento apresentando uma classificação de técnicas de correspondência e algoritmos de similaridade para a comparação de conceitos. Por fim, a Seção 2.3 aborda sobre dados complexos e consultas por similaridade.

2.1 Representação de Conhecimento

A representação de conhecimento é um ramo da Inteligência Artificial (IA) que consiste em formalizar o conhecimento explicitado para que seja interpretado e utilizado por sistemas computacionais. Esses sistemas possuem um modelo computacional de algum domínio de interesse no qual os símbolos representam artefatos do mundo real, como conceitos, objetos, eventos etc. Um sistema baseado em conhecimento mantém uma base de conhecimento que armazena os símbolos do modelo computacional na forma de declarações sobre o domínio. As declarações dispõem de alguma linguagem de representação para descrever formalmente o conhecimento armazenado por meio de diagramas, modelos e textos (WIELINGA, 2013; SOWA, 2000; GRIMM; HITZLER; ABECKER, 2007).

As linguagens de representação de conhecimento disponibilizam primitivas que buscam capturar e estruturar conceitos de um determinado domínio, ao mesmo tempo em que retém sua representatividade semântica. São baseadas em diferentes técnicas ou métodos de representação, como Redes Semânticas, regras, lógica de primeira ordem, MC, Ontologias, entre outras. Também é possível usar combinações dessas técnicas (Sistemas Híbridos) (ELMASRI; NAVATHE, 2011; ABEL; FIORINI, 2013).

Regra, no âmbito de IA, é o tipo mais usual de representação de conhecimento. Pode ser definido como uma estrutura IF-THEN aplicada a uma proposição, que relaciona informações ou fatos da parte IF <condição> a alguma ação na parte THEN <conclusões e

ações>. As regras fornecem uma descrição de como resolver um problema e são relativamente fáceis de criar e entender (NEGHEVITSKY, 2005).

Uma proposição, em termos lógicos, é uma sentença declarativa para a qual pode ser atribuído um valor verdadeiro ou falso (PRIEST, 2001). Na representação de conhecimento uma proposição tem um significado subjacente, representando uma relação entre conceitos. Minimamente, uma proposição é composta de um sujeito e um predicado (STERNBERG; STERNBERG, 2011).

Como exemplo a frase, “Penso, logo existo” é uma proposição causal composta da regra com estrutura IF-THEN formalizada com a lógica de primeira ordem, a qual mantém uma relação semântica de causa e efeito. As relações de causa e efeito são decisivas na identificação de eventos e processos importantes. Estão presentes em sistemas dinâmicos e complexos, bem como no pensamento racional (CAO; SUN; ZHUGE, 2018).

Dentre as técnicas de representação de conhecimento citadas, existem as que proporcionam a representação gráfica das estruturas do conhecimento. Esse tipo de representação permite que atalhos sejam tomados no processo cognitivo e na análise da estrutura do conhecimento (CARD; MACKINLAY; SHNEIDERMAN, 1999).

Considerando o escopo do trabalho, serão consideradas as principais técnicas utilizadas na comparação de MC e das representações gráficas do conhecimento que mais se assemelhem à sua estrutura, ou seja, que organizem o conhecimento na forma de proposições com relações de causa e efeito. Os MC, Redes Semânticas e Ontologias se enquadram nesse contexto e serão abordados a seguir.

2.1.1 Mapas Conceituais

Os MC surgiram em 1972 desenvolvidos em um estudo de Joseph D. Novak, que visava através de entrevistas periódicas entender o comportamento de crianças previamente instruídas com lições de ciência e crianças não instruídas (NOVAK; MUSONDA, 1991). O estudo de Novak, foi baseado na psicologia de aprendizagem significativa de David Ausubel (1963), que como ideia central diz que a aprendizagem é um processo de assimilação de novos

conceitos que formam estruturas proposicionais de um aprendiz (NOVAK; CAÑAS, 2008). A aprendizagem é dita significativa quando uma nova informação (conceito, ideia, proposição) adquirida pelo aprendiz, relaciona-se com as ideias, conceitos e proposições já existentes em sua estrutura relevante de conhecimentos (MOREIRA, 1998). Toda essa estrutura de conhecimento é conhecida por estrutura cognitiva do indivíduo (NOVAK; CAÑAS, 2008). O MC surgiu como uma ideia para melhor representar a compreensão conceitual das crianças e com isso nasceu uma ferramenta para uso em pesquisas e em outras aplicações (NOVAK; CAÑAS, 2008).

A teoria do conceito é um campo extremamente amplo, interdisciplinar e complexo (HJORDLAND, 2009) que possui diferentes posicionamentos filosóficos e científicos, dificultando sua definição (MELO; BRÄSCHER, 2015). Segundo o filósofo grego Aristóteles (384-322 a.C.) apresentado por Abbagnano (2003), os conceitos se diferem das palavras e das coisas por terem uma realidade mental; o conceito é o modo como os homens organizam mentalmente todas as coisas existentes. O contexto é de grande importância para a seleção do significado do conceito (MURPHY, 2002), podendo basear-se nas interações sociais de uma comunidade (CLARK, 1996) ou na relação entre os conceitos em uma declaração (CRUSE, 1986).

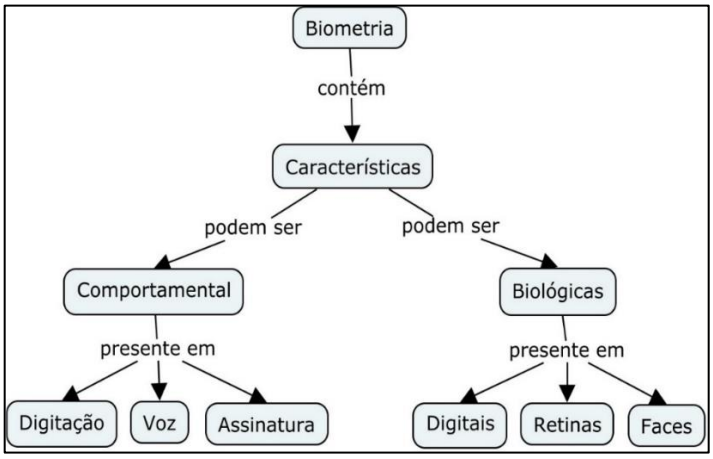
No âmbito de MC, os conceitos são elementos da representação temática da informação, responsáveis por descrever as propriedades de um objeto ou construir um enunciado lógico e verdadeiro sobre eventos ou situações que possuem atributos comuns em um contexto definido (HJORLAND, 2009; PÉREZ; VIEIRA, 2005). O relacionamento entre conceitos pode ser estático ou dinâmico (SAFAYENI, DERBENTSEVA; CAÑAS, 2005). A relação estática descreve, define, organiza o conhecimento e está presente na atual forma dos MC. A relação dinâmica, para dois conceitos quaisquer, atua em como a mudança de um conceito afeta o outro conceito, unitária ou coletivamente (SAFAYENI; DERBENTSEVA; CAÑAS, 2005). O ato de definir e relacionar os conceitos indica aquisição de conhecimento, produto de uma aprendizagem significativa por parte do aprendiz (MOREIRA, 1987).

Os MC podem ser definidos como ferramentas gráficas que são utilizados para organizar e representar o conhecimento da estrutura cognitiva de um indivíduo, através de diagramas (NOVAK; CAÑAS, 2008). A estruturação do MC se apoia em uma pergunta prévia, denominada de questão focal, da qual originam todos os conceitos. Os conceitos são as palavras

contidas em círculos ou retângulos, como mostra a Figura 2.1. Para vincular dois conceitos e evidenciar o motivo de seu relacionamento, é usada uma linha contendo um termo de ligação que pode ser um verbo ou sintagma verbal. Esses conceitos mantém uma relação de causa e efeito e, juntamente com o termo de ligação, formam uma proposição (NOVAK; CAÑAS, 2008; NOVAK, 2010).

A forma estrutural mais comum do MC é o do tipo hierárquico e Novak e Cañas (2008) enfatizam que os conceitos são representados seguindo uma estrutura hierárquica lógica, onde os conceitos mais abrangentes são alocados no topo dos mapas e quanto mais específicos, menos gerais, organizados hierarquicamente abaixo como na Figura 2.1.

Figura 2.1 - Mapa Conceitual hierárquico



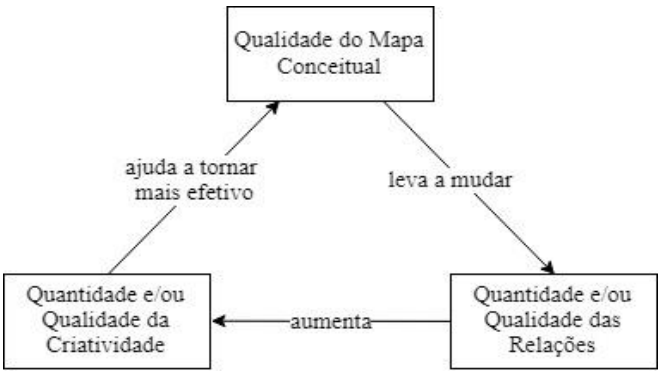
Fonte: Desenvolvido pelo autor

Analisando Mapas Conceituais metodologicamente e conceitualmente, não há uma obrigatoriedade de representá-los em uma estrutura hierárquica. Porém, se o mapa estiver estruturado de forma hierárquica, convém observá-lo de acordo com essa estrutura (RUIZ-PRIMO; SHAVELSON, 1996).

Processos não lineares são representados nos Mapas Conceituais Cíclicos (Figura 2.2), ou diagramas de *feedbacks*, que possuem uma estrutura onde o conceito-causa é retroalimentado pelo conceito-efeito, mesmo que indiretamente. Nesse processo, os conceitos são altamente interdependentes, ou seja, qualquer mudança no estado de qualquer conceito será responsável por afetar todos os outros conceitos do mapa (SAFAYENI; DERBENTSEVA; CAÑAS, 2005; SENGE, 2014; SAFAYENI; DERBENTSEVA; CAÑAS, 2006).

A Figura 2.2 demonstra um Mapa Conceitual cíclico contendo três conceitos e três proposições. A alteração no estado de qualquer um dos três conceitos relacionados, afetará positivamente ou negativamente os outros dois conceitos, demonstrando a interdependência dos conceitos no Mapa Conceitual cíclico.

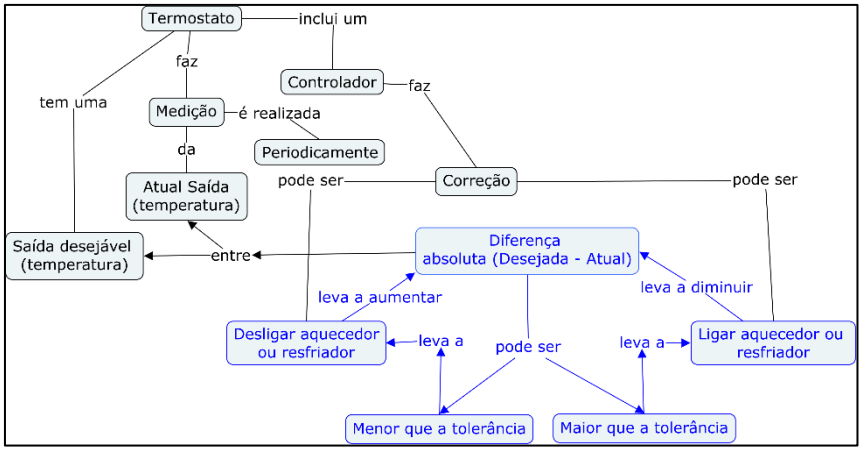
Figura 2.2 - Mapa Conceitual cíclico



Fonte: adaptado de Safayeni, Derbentseva e Cañas (2005)

Um mapa híbrido faz a união entre duas topologias de mapas. A Figura 2.3 mostra um mapa híbrido de um termostato que contém relações lineares, específicas de um Mapa Conceitual hierárquico e retroalimentações, como em um Mapa Conceitual cíclico.

Figura 2.3 – Estrutura de um mapa híbrido



Fonte: adaptado de Safayeni, Derbentseva e Cañas (2005)

Nesse caso, o Mapa Conceitual híbrido foi utilizado para fornecer mais informações sobre a operação do sistema (Mapa Conceitual hierárquico) e descrever o processo com mais precisão, mostrando que é contínuo, cíclico e dinâmico (Mapa Conceitual cíclico) (SAFAYENI; DERBENTSEVA; CAÑAS, 2005). Não existe um mapa MC “correto”, mas sim

um mapa conceitual para um determinado conteúdo segundo os significados que uma pessoa atribui aos conceitos e às relações entre eles (MOREIRA, 1998).

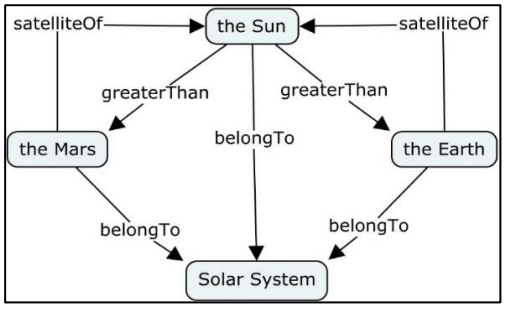
2.1.2 Redes Semânticas

As Redes Semânticas são estruturas utilizadas para a representação de qualquer tipo de conhecimento que possa ser descrito em linguagem natural (LEHMANN, 1992). Na inteligência artificial, as Redes Semânticas foram introduzidas a partir do estudo de Quillian (1968) sobre a organização da memória semântica humana e são definidas como um formato capaz de armazenar palavras e seus significados.

A estrutura mais usual de uma Rede Semântica é um grafo dirigido, cuja forma é composta por uma coleção de nós (vértices) conectados por arcos (arestas) que expressam as ligações entre os nós (MARKOWITZ; NUTTER; EVENS, 1992). Os nós são rotulados por textos descritivos que podem representar conceitos ou objetos que compõem um domínio do conhecimento. Os arcos definem relações entre nós e seus rótulos especificam o tipo da relação entre eles (LEHMANN,1992; MARINOV; ZHELIAZKOVA, 2005).

A *Semantic Link Network* (SLN) é um tipo de Rede Semântica relacional auto-organizada, criada para representar conhecimento básico, porém, preservando a riqueza semântica. Sua estrutura consiste em nós semânticos, de representação bastante abrangente (e.g., conceitos, eventos, textos, imagens ou até uma SLN), conectados por links semânticos. O *link* semântico entre os nós representa um tipo de conhecimento relacional, composto por um ponteiro com uma *tag* que descreve as interações semânticas como *cause-effect*, *implication*, *subtyp*e, etc. A semântica das tags é regulada por sua categoria, regras de raciocínio e casos de uso (ZHUGE, 2012; ZHUGE, 2009). A Figura 2.4 mostra um simples exemplo estrutural de uma SLN, representando um pequeno trecho relacional do sistema solar, contendo sete links semânticos e quatro conceitos.

Figura 2.4 – Estrutura da SLN do sistema solar



Fonte: adaptado de Zhang e Sun (2010)

Esse tipo de Rede Semântica é um modelo semântico geral, que serve para modelar a estrutura e evolução de sistemas complexos, composto por relações de causa-efeito e retroalimentações, importantes para representação e compreensão desses sistemas (CAO; SUN; ZHUGE, 2018).

2.1.3 Ontologias

As Ontologias podem ser definidas como estruturas de representação de conhecimento baseadas na especificação explícita de uma conceitualização (GRUBER, 1995). A conceitualização é um conjunto de conceitos, objetos e outras entidades sobre um conhecimento que está sendo representado e o relacionamento que se mantém entre eles. A especificação refere-se aos termos de linguagem e vocabulário usados para especificar a conceitualização (ELMASRI; NAVATHE, 2011; NOY; HAFNER, 1997).

O estudo das Ontologias visa descrever genericamente as estruturas e os relacionamentos possíveis na realidade ou em um domínio de interesse por meio de um vocabulário comum (ELMASRI; NAVATHE, 2011).

Tecnicamente, os principais constituintes de uma Ontologia são conceitos, relações e instâncias. Os conceitos são representados como nós genéricos presentes em Redes Semânticas e MC. Eles representam as categorias ontológicas, relevantes no domínio de interesse. As relações se comportam como arcos das Redes Semânticas e linhas de MC, conectando os conceitos semanticamente e especificando suas inter-relações. Instâncias são mapeadas para nós individuais em Redes Semânticas e representam os objetos concretos nomeados e identificáveis no domínio de interesse (GRIMM; HITZLER; ABECKER, 2007).

Quando projetadas computacionalmente, as Ontologias podem ser visualizadas graficamente ou de modo formal, baseada nas técnicas de representação de conhecimento discutidas anteriormente. Sob a perspectiva de um engenheiro do conhecimento, uma Ontologia é frequentemente apresentada em forma de Rede Semântica (GRIMM; HITZLER; ABECKER, 2007).

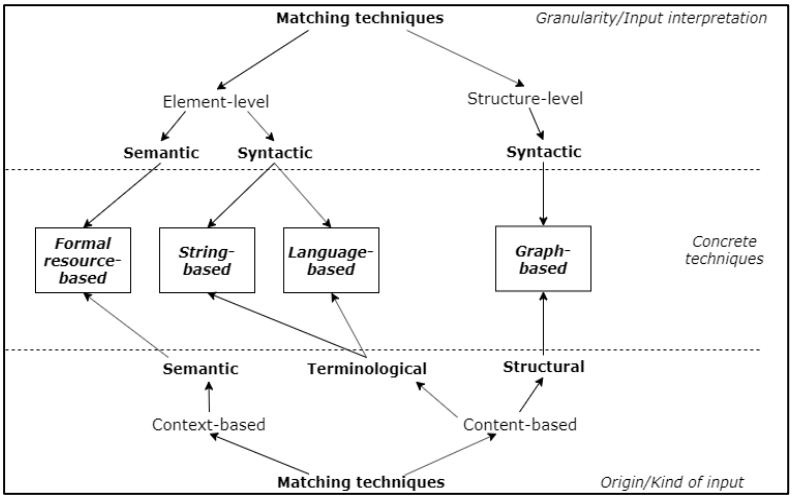
2.2 Comparação de Conhecimento

Os modelos de representação de conhecimento são comprovadamente úteis para a organização e visualização do conhecimento explícito. Porém, para consultar e comparar computacionalmente as representações resultantes desses modelos, é necessário o uso de alguma técnica que compare suas características.

A comparação de conhecimento representado por MC, Redes Semânticas e Ontologias é um processo complexo que considera as características estruturais desses modelos, sob diferentes tipos de análise. Entretanto, a solução normalmente envolve alguma técnica de correspondência.

Técnicas de correspondência abrangem diferentes medidas de similaridade, estratégias e metodologias de comparação, que podem ser usadas em modelos e sistemas. Processos de comparação são propostos em trabalhos que, em sua maioria, não se dedicam a uma única técnica de correspondência, além de poderem ser aplicados a mais de um modelo de representação de conhecimento (OTERO-CERDEIRA; RODRÍGUEZ-MARTÍNEZ; GÓMEZ-RODRÍGUEZ, 2015). Desse modo, a revisão e classificação desses trabalhos não é uma tarefa trivial. As classificações das técnicas de correspondência básicas (OTERO-CERDEIRA; RODRÍGUEZ-MARTÍNEZ; GÓMEZ-RODRÍGUEZ, 2015) consideradas estão divididas em categorias e apresentadas na Figura 2.5.

Figura 2.5 – Classificação das técnicas de correspondência



Fonte: adaptado de Otero-Cerdeira, Rodríguez-Martínez e Gómez-Rodríguez (2015)

Considerando as categorias presentes na Figura 2.5 e seguindo as descrições presentes em (OTERO-CERDEIRA; RODRÍGUEZ-MARTÍNEZ; GÓMEZ-RODRÍGUEZ, 2015; MARSHALL; CHEN; MADHUSUDAN, 2006), a classificação se inicia de cima para baixo e tem foco na interpretação que as diferentes técnicas oferecem nas informações de entrada. Os elementos do primeiro nível podem ser classificados como:

- *Element-level matchers*: técnicas dessa categoria realizam correspondências a partir dos conceitos presentes nos modelos de representação de conhecimento isoladamente, ignorando a estrutura;

- *Structure-level* matchers: são técnicas que realizam correspondências analisando como os conceitos se relacionam na estrutura.

No segundo nível de classificação, tem-se:

- *Semantic*: são técnicas que utilizam semântica formal para interpretar os dados de entrada e justificar os resultados obtidos;

- *Syntactic*: são técnicas que limitam sua interpretação de entrada com base em instruções declaradas nos algoritmos correspondentes.

Ainda considerando a Figura 2.5, a classificação das técnicas de correspondência pode também ser realizada de baixo para cima, podendo ser inicialmente dividida em duas categorias, determinadas pela origem das informações obtidas. O primeiro nível de classificação contém as seguintes categorias:

- *Content-based*: são técnicas que focam nas informações internas dos modelos de representação para realizar a correspondência;

- *Context-based*: são técnicas que consideram informações externas que podem ser advindas de relações entre cada modelo ou outros recursos externos (contexto).

O segundo nível de classificação refina e divide as duas categorias. A categoria *Content-based* é dividida em dois grupos, dependendo da entrada que as técnicas utilizam:

- *Terminological*: técnicas que consideram suas entradas como Strings;

- *Structural*: utilizam técnicas que se baseiam na estrutura das entidades (conceitos, objetos, eventos e relações) em cada modelo de representação de conhecimento.

A categoria *Context-based* agrupa técnicas classificadas como semânticas (*Semantic*).

O último nível contempla qualquer uma das classificações apresentadas e corresponde a técnicas específicas. Os trabalhos pesquisados serão classificados de acordo com a metodologia apresentada, podendo pertencer a mais de uma categoria de correspondência, como a seguir:

- *Formal Resource-based*: essas técnicas utilizam de algum recurso formal para apoiar o processo de correspondência, como Ontologias de nível superior (MASCARDI; LOCORO; ROSSO, 2009) e Redes Semânticas SNePS (SAPHIRO, 2000);

- *String-based*: são técnicas que verificam a similaridade sintática entre Strings que representam os conceitos presentes em Ontologias, MC e Redes Semânticas. Existem diversos

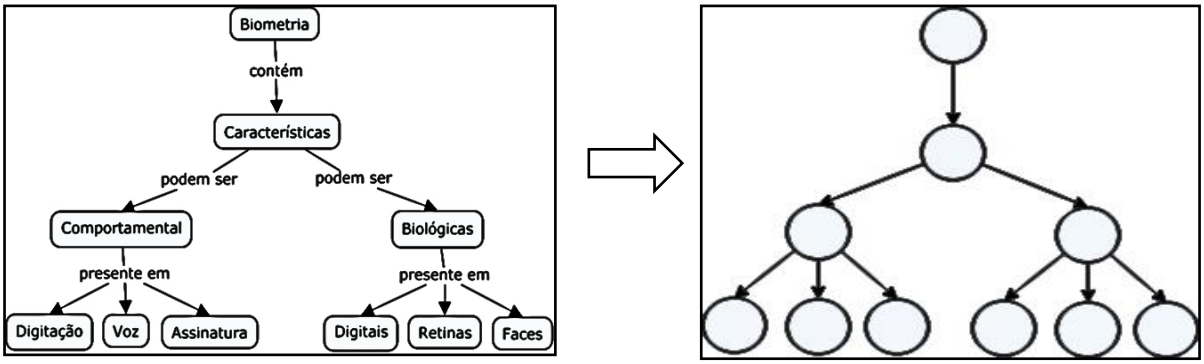
algoritmos que realizam essa verificação como *Jaro*, *TFIDF*, *Jaro-Winkler*, *Leveshtein* etc. (COHEN; RAVIKUMAR; FIENBERG, 2003; TAN; STEINBACH; KUMAR, 2006). A aplicabilidade desses algoritmos pode ser vista em: (COHEN; RAVIKUMAR; FIENBERG, 2003; ANGELES; ESPINO-GAMEZ, 2015), onde são comparados, combinados e avaliados em termos de precisão, por meio de variações de determinados parâmetros e tempo de execução.

- *Language-based*: são técnicas baseadas no processamento de linguagem natural e na semântica dos conceitos, considerando *Strings* como palavras em alguma linguagem natural. Remoção de *stop words* e *stemming* são algumas das técnicas aplicadas por (SHAH; SYEDA-MAHMOOD, 2004). As *stop words* são palavras como pronomes, preposições e advérbios que não acrescentam significado a uma String e por isso são removidas. O processo de *stemming* é responsável por identificar variações morfossintáticas de cada palavra, visando reduzir palavras flexionadas ou derivadas ao seu radical (*stem*). Esta categoria também considera o uso de sinônimos para identificar relações semânticas entre palavras utilizando algum recurso externo (*dictionary*, *lexicons* ou *thesauri*). O trabalho de He, Yang e Huang (2001), por exemplo, utiliza o banco de dados *WordNet*[1]como um recurso externo.

- *Graph-based*: devido à semelhança estrutural de Redes Semânticas, Ontologias e MC, é possível estabelecer uma equivalência de cada modelo com a estrutura de um grafo (Figura 2.6). Por exemplo, um MC pode ser convertido em um grafo de atributos $G = (V, E)$, onde V corresponde aos conceitos e E às arestas que representam cada relação entre dois conceitos. As arestas do grafo podem conter frases ou palavras para descrever a relação entre os conceitos (DE SOUZA *et al*., 2008). Diante disso, essa categoria considera os modelos de representação de conhecimento como grafos, ou mesmo árvores, e em alguns trabalhos, tratam o problema de correspondência como um problema de isomorfismo ou homomorfismo de grafos. Exemplos dessas e outras técnicas podem ser encontrados em (DE SOUZA *et al*., 2008; LIMONGELLI *et al.*, 2016; MONTES-Y-GÓMEZ; LÓPEZ-LÓPEZ; GELBUKH, 2000; JOSLYN; PAULSON; WHITE, 2009).

[1] Disponível em: http://wordnet.princeton.edu/

Figura 2.6 - MC transformado em grafo



Fonte: Desenvolvido pelo autor

As técnicas apresentadas são a base constitutiva das *técnicas de correspondência complexas*, que propõem a utilização de maneiras diferentes de tratar o problema de correspondência, combinando técnicas de diferentes categorias (OTERO-CERDEIRA; RODRÍGUEZ-MARTÍNEZ; GÓMEZ-RODRÍGUEZ, 2015). Alguns exemplos podem ser observados em (COHEN; RAVIKUMAR; FIENBERG, 2003), onde os autores combinam diferentes medidas de similaridade para gerar uma função híbrida. No trabalho de De Souza *et al.* (2008), os autores combinam o uso de isomorfismo para comparar os grafos de MC com técnicas da categoria *Language-based* para tratar a semântica de seus conceitos, a partir da identificação de similaridades.

A análise feita nesta seção apresentou trabalhos e categorias aplicáveis ou adaptáveis a MC, Redes Semânticas e Ontologias, em razão da semelhança estrutural entre esses modelos de representação de conhecimento. É possível concluir que apenas o trabalho (MASCARDI; LOCORO; ROSSO, 2009) da categoria *Formal resource-based* não pode ser utilizado para todos os modelos, por conter técnicas específicas para um tipo de Ontologia.

É possível observar que existem vários tipos de técnicas de correspondência, tanto básicas como complexas, que utilizam o critério de similaridade. E como constatado em (OTERO-CERDEIRA; RODRÍGUEZ-MARTÍNEZ; GÓMEZ-RODRÍGUEZ, 2015), muitas outras continuarão a surgir. Algumas delas poderão ser versões revisadas de técnicas já existentes, mas a tendência é que novos tipos de técnicas sejam elaboradas, especialmente para categorias e modelos ainda não explorados totalmente como o MCE.

Comparações aplicadas especificamente para conceitos contam com algoritmos sintáticos e semânticos difundidos na literatura. As próximas seções apresentam uma análise mais ampla sobre conceitos e algoritmos que podem ser utilizados na comparação de conceitos.

2.2.1 Comparação de conceitos

Não existem regras para limitar o que usar como conceito, visto que, cada pessoa tem seu modo singular de entender alguma coisa. Porém, existem boas práticas e limitações impostas na utilização de conceitos que facilitam a construção e o entendimento de MC. Frases longas para a representação de conceitos, conceitos desconexos com o contexto do mapa, mais de 20 conceitos (NOVAK; CAÑAS, 2008) são exemplos que estão fora dos padrões.

Como explanado na Seção 2.1, os MC são compostos de uma estrutura de relação entre conceitos, que dão sentido a um contexto previamente definido. Sabendo que conceitos refletem conhecimento e são rotulados por uma ou mais palavras, a sintática e a semântica aplicada a um conceito por um agente do conhecimento é um fator preponderante para sua comparação.

Para exemplificar, um Mapa Conceitual pode conter uma proposição “pessoas – fumam – cigarro” enquanto em outro a proposição pode estar representada como “cidadãos – pitam – tabaco”. Apesar das duas proposições conterem palavras diferentes, elas são sinônimas e semanticamente iguais. Além do problema de palavras sinônimas, outros problemas semânticos podem ser encontrados (Tabela 2.1) e com isso algoritmos que consideram apenas a parte sintática de palavras são incompletos.

Tabela 2.1 – Problemas semânticos da língua portuguesa

Problemas semânticos	**Descrição**	**Exemplo**
Sinônimos	Palavras diferentes que contém o mesmo significado	Cão – Cachorro Cigarro – Tabaco
Homônimos	Palavras iguais que contém significados diferentes	Verão (substantivo) – Verão (verbo) Leve (substantivo) – Leve (verbo)
Parônimos	Palavras com escrita e pronúncia semelhantes, mas com significados diferentes	Soar (emitir som) – Suar (transpirar) Cumprimento (cumprir) – Comprimento (medida)
Polissemia	Palavras que contém diversos significados de acordo com o contexto no qual está inserida	Cabo (cabo de vassoura, posto militar) Manga (fruta, parte da roupa)

Fonte: Desenvolvido pelo autor

O problema de semelhança sintática e semântica de conceitos pode ser tratado com a utilização da técnica de similaridade. Por meio de uma função de similaridade ou dissimilaridade é possível descrever o grau de similaridade entre dois elementos, comparando suas características semelhantes e diferentes (IFENTHALER, 2008). A representação do valor de similaridade está entre 0 e 1, sendo que se dois conceitos não compartilharem características em comum, então são completamente diferentes e possuem valor de similaridade 0. No caso de possuírem todas as características em comum, o valor de similaridade será 1 (IFENTHALER, 2008).

2.2.2 Algoritmos para comparação de conceitos

Para a comparação de conceitos existem diversos algoritmos de similaridade que se limitam a condições específicas e variam seu nível de complexidade. Diante disso, os seguintes algoritmos de similaridade são descritos de acordo com seu tipo.

Algoritmos baseados em caracteres

- *Levenshtein*: utilizada para comparar *Strings* através de operações de substituição, remoção e inserção de caracteres. As operações visam definir quantas modificações são necessárias para transformar uma *String* em outra. A distância entre duas *Strings* é definida como o número de operações necessárias para sua combinação. Uma pontuação de 0 implica que nenhuma operação foi necessária, ou seja, as *Strings* são iguais (FREEMAN; CONDON; ACKERMAN, 2006).

- *Smith Waterman*: baseado no algoritmo de Levenshtein, contendo pequenas modificações. Esse algoritmo alinha duas *substrings* e qualquer pontuação negativa é substituída por zero, essa pontuação passa a ser a melhor dentre todas. Com isso, o começo e o fim das duas *Strings* não precisam estar alinhados (SMITH; WATERMAN, 1981).

- *Jaro*: compara duas *Strings* baseada no número e ordem de seus caracteres em comum e visa trabalhar com *Strings* pequenas (COHEN; RAVIKUMAR; FIENBERG, 2003). Considerando duas *Strings* (*s* e *t*) a similaridade de Jaro (1995) é:

$$sim_j = \frac{m}{3a} + \frac{m}{3b} + \frac{m}{3a} + \frac{m-t}{3m}$$

Onde, *m* é o número de correlações de caracteres, *a* e *b* são os tamanhos de *s* e *t* respectivamente e *t* é o número de transposições.

- *Jaro-Winkler*: é a extensão de Jaro que fornece um peso maior a cadeias de caracteres que coincidam desde o início, para um tamanho de prefixo estabelecido (MARTINS; MACHADO, 2018). A similaridade de Jaro-Winkler é:

$$sim_{jw} = sim_j + lp(1 - sim_j)$$

Onde, sim_j é a similaridade de Jaro, *l* é o comprimento do prefixo comum no início da cadeia até um máximo de 4 caracteres, *p* é um fator de escala que atribui pontuação para os prefixos. A métrica de comparação é 0 quando as *Strings* forem totalmente diferentes e 1 quando forem iguais (WINKLER, 2006). A distância é dada por: $d_w = 1 - sim_{jw}$.

Algoritmos baseados em termos ou *tokens* e algoritmos híbridos

- *Jaccard*: é uma medida que calcula a similaridade entre dois vetores binários, é similar à similaridade cosseno pois não considera as combinações 0-0, a medida é aplicada em multiconjuntos (*bag-of-words*) e palavras (*tokens*) (TAN; STEINBACH; KUMAR, 2006; COHEN; RAVIKUMAR; FIENBERG, 2003). Dado um conjunto de palavras *(S* e *T*), então a similaridade é dada por: $J = \frac{|S \cap T|}{|S \cup T|}$ (COHEN; RAVIKUMAR; FIENBERG, 2003).

- *Monge-Elkan*: esse algoritmo híbrido simples e eficaz mede a similaridade entre duas longas *strings* contendo múltiplos tokens, utilizando uma similaridade interna *sim*(*a, b*) capaz de medir a similaridade dos tokens *a* e *b* individualmente (ANGELES; ESPINO-GAMEZ, 2015). Dado dois textos *A* e *B*, sendo seus respectivos números de tokens |*A*| e |*B*| o algoritmo calcula a média dos valores de similaridade entre os pares de tokens mais semelhantes presentes nos textos *A* e *B* (ANGELES; ESPINO-GAMEZ, 2015). Sua fórmula consiste em:

$$MonElk = \frac{1}{|A|} \sum_{i=1}^{|A|} \max_{j=1}^{|B|} \{sim'(A_i, B_j)\}$$

- *Soft TF-IDF*: Definido por Cohen, Ravikumar e Fienber (2003), esse algoritmo híbrido trabalha com o TFIDF ou Similaridade do cosseno em uma versão modificada ("soft"). TFIDF é definido como:

$$TFIDF(S,T) = \sum_{w \in S \cap T} V(w,S) \times V(w,T)$$

Onde *S* e *T* são dois conjuntos de palavras (*strings*) ou tokens quaisquer, $TF_{w,S}$ é a frequência da palavra *w* em S, *N* o tamanho do "corpus", IDF_w é o inverso da fração de nomes no corpus que contém *w*, $V'(w,S) = \log\left(TF_{w,S} + 1\right) \times \log(IDF_w)$ e $V(w,S) = V'(w,S) \,/\, \sqrt{\sum_{w'} V'(w,S)^2}$ (COHEN; RAVIKUMAR; FIENBERG, 2003).

Na versão "soft" os autores estabelecem que tokens similares são considerados como tokens em $S \cap T$, *dist* é uma segunda função de similaridade (autores utilizam Jaro-Winkler), (*CLOSE*(θ,*S*,*T*) é o conjunto de palavras $w \in S$, tais que, existe algum $v \in T$ tal que $dist'(w,v) > \theta$, e para $w \in CLOSE(\theta,S,T)$, seja $D(w,T) = max_{v \in T}\, dist(w,v)$. É definido que:

$$SoftTFIDF(S,T) = \sum_{w \in CLOSE(\theta,S,T)} V(w,S) \times V(w,T) \times D(w,T)$$

Os algoritmos apresentados tratam de comparações sintáticas, mas o objetivo é que essas comparações aufiram resultados relevantes para comparações semânticas.

2.3 Dados complexos e consultas por similaridade

Muitas aplicações informatizadas requerem o armazenamento e manipulação de novos tipos de dados, denominados complexos. Dados complexos são aqueles cuja estrutura interna é composta por vários atributos mais simples, mesmo que essa estrutura não seja reconhecida pelo Sistema de Gerenciamento de Bancos de Dados (SGBD). Podem ser classificados em três tipos principais considerando SGBDs relacionais: estruturados, semiestruturados e não estruturados (ELMASRI; NAVATHE, 2011; SILBERCHATZ; KORTH; SUDARSHAN, 2019).

Um dado complexo estruturado tem sua estrutura definida e conhecida pelo SGBD. Eles permitem que os atributos (de valores múltiplos, compostos ou atômicos) sejam representados diretamente por meio do Modelo Entidade-Relacionamento (MER). O dado complexo semiestruturado é uma forma de dado estruturado que não obedece a estrutura formal de modelos de dados associados a bancos de dados relacionais ou outras formas de tabelas de dados. Contém tags ou outros marcadores para descrever elementos semânticos e impor hierarquias de registros e campos de dados. É conhecido como estrutura autodescritiva (BUNEMAN, 1997). Um dado complexo não estruturado é um tipo de dado que SGBD não sabe qual é sua estrutura, apenas a aplicação que os utiliza pode interpretar seu significado (ELMASRI; NAVATHE, 2011; SILBERCHATZ; KORTH; SUDARSHAN, 2019). Na maioria das vezes requer grandes quantidades de bytes de memória para armazenamento, como uma imagem ou um texto longo.

Para serem úteis em sistemas computacionais, não basta armazenar dados complexos, mas também se faz necessário comparar o seu conteúdo, a fim de realizar consultas e análises. De modo geral, a comparação de dois dados complexos é realizada a partir das características de sua estrutura interna, cuja técnica denomina-se extração de características. As características devem ser extraídas do dado complexo de acordo com o domínio em que esteja inserido. Desse modo, o dado complexo é descrito por um vetor de características (MULLER *et al.*, 2004).

Domínios de dados complexos podem ser representados por dois modelos principais (SAMET, 2006): modelo de espaço vetorial e modelo de espaço métrico. No modelo de espaço vetorial os dados complexos são descritos por vetores de características, tratados como coordenadas de pontos no espaço *e*-dimensional, onde *e* corresponde à quantidade de elementos (características) que compõem o vetor de características. Por sua vez, para alguns domínios, a extração de vetores de características com a mesma dimensão para todos os dados complexos pode ser uma tarefa muito complicada, ou até inviável. Ou seja, o número de características varia para cada dado complexo, não havendo dimensão definida. Nesse caso, os dados complexos são adimensionais, o que define o modelo de espaço métrico.

De modo geral, para comparações de dados complexos utiliza-se o critério de similaridade, que busca a semelhança de dois dados complexos por meio de suas características (SAMET, 2006; CHAVÉZ et al., 2001). Assim, a partir do armazenamento de dados complexos

em um banco de dados, suas características podem ser utilizadas nas operações de comparação efetuadas para responder consultas por similaridade.

Consultas por similaridade são consultas que identificam o grau de similaridade entre dados complexos e envolvem: um objeto de busca, também considerado como o centro da consulta, que é o dado a partir do qual se deseja encontrar os mais similares; uma medida de similaridade; e um conjunto de parâmetros dependentes do tipo de consulta por similaridade a ser realizada (SAMET, 2006).

Os tipos mais usuais de consultas por similaridade são (SAMET, 2006):

- Consulta por Abrangência (*Range Query – RQ*): recupera todos os objetos similares a um objeto central de busca s_q dentro de um valor estabelecido r_q (raio de busca).

- Consulta aos *k*-vizinhos mais próximos (*k-Nearest Neighbors Query - kNNQ*): recupera os *k* objetos mais similares a um objeto central de busca s_q.

Segundo Chávez *et al*. (2001) e Samet (2006), as consultas por similaridade são apoiadas por estruturas de dados para espaços métricos, que englobam os dois modelos citados anteriormente (espaço vetorial e espaço métrico). O espaço métrico formado por um par $\mathsf{M} = (\mathsf{S}, d)$ onde S é um domínio ou universo de objetos válidos e $d(\ )$ é uma métrica que corresponde à medida de distância entre dois objetos, quanto menor a distância entre eles, mais próximos ou semelhantes eles serão. Para validar a métrica, ela deve satisfazer as propriedades de simetria $\forall\, x, y \in \mathsf{S},\ d(x, y) = d(y, x)$, não-negatividade $\forall\, x, y \in \mathsf{S},\ d(x, y) > 0$ e $d(x, x) = 0$ e desigualdade triangular $\forall\, x, y, z \in \mathsf{S},\ d(x, y) \leq d(x, z) + d(z, y)$.

O espaço vetorial pode ser considerado um caso particular de espaço métrico, onde os elementos *e* do vetor de características são representados por *e* coordenadas de valores reais, nesse caso, as métricas mais comuns são as da família L_p (Minkowski): Distância de bloco ou Manhattan, Distância Euclidiana e Infinity (SAMET, 2006).

Em suma, a avaliação de similaridade entre dois dados complexos de um mesmo domínio é realizada a partir da comparação de suas características, por meio de funções matemáticas ou mesmo algoritmos computacionais, de modo a se obter o grau de similaridade

entre eles. Usualmente, o grau de similaridade é um valor entre 0 e 1, onde 1 significa que os dados complexos comparados são totalmente iguais e 0 totalmente diferentes.

2.4 Considerações finais do capítulo

A análise feita no capítulo apresentou trabalhos e categorias aplicáveis ou adaptáveis a MC, Redes Semânticas e Ontologias, em razão da semelhança estrutural entre esses modelos de representação de conhecimento. Foram abordados os principais algoritmos de comparação por similaridade para a comparação de conceitos. Por fim, a fundamentação teórica de dados complexos e consultas por similaridade foi apresentada. No próximo capítulo serão abordados MCE.

3 MAPAS CONCEITUAIS ESTENDIDOS

O presente capítulo tem como objetivo apresentar objetivamente o MCE, apontando suas características e aplicações.

3.1 Apresentação e características do MCE

O MCE é um modelo de representação de conhecimento, formado por regras baseadas na abordagem proposicional. Esse modelo é baseado em MC e é utilizado para representar um modelo mental individual, revelando sua estrutura complexa para que possa ser analisada e compreendida (ZAMBON *et al*., 2016a). Sua concepção parte da teoria de memória de trabalho de Baddeley (1992), e técnicas do âmbito da gestão estratégica e da gestão do conhecimento.

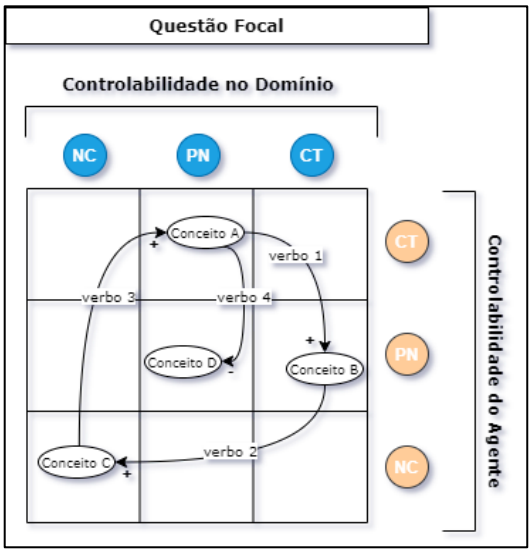
Os modelos mentais são o conhecimento tácito de indivíduos e contam com estruturas complexas que relacionam questões naturais (sentidos, capacidades individuais), ambientais (experiências) e culturais (relacionadas ao lugar, à época, entre outros). Esses modelos além de descrever conceitos e suas relações, como em um mapa conceitual, também os posicionam em uma matriz de atributos, que representa as impressões particulares do indivíduo sobre os conceitos e sobre o cenário onde estão contidos (ZAMBON *et al*., 2016a; JARAMILLO, 2018).

Assim como nos MC, a modelagem de um MCE se inicia pela declaração de uma questão focal, com a qual se define um problema ou questão a ser explicada. A partir da questão focal é criada uma estrutura de conceitos relacionados em causa e efeito por verbos e pesos. A estrutura pode ser hierárquica, cíclica ou híbrida, e estará contida em uma matriz de atributos com três linhas e três colunas. A matriz representa a controlabilidade do indivíduo sobre um domínio conceitual (ZAMBON *et al*., 2016a). Para exemplificar, a modelagem de um MCE é representada na Figura 3.1.

Os conceitos possuem significado intrínseco e também contextualizado pelos atributos da matriz em que estão inseridos. Cada conceito se associa à questão focal por

uma escala de forças crescente da esquerda para a direita e de baixo para cima, na seguinte ordem: não controlável (*NC*), penumbra (*PN*) e controlável (*CT*) (ZAMBON *et al.*, 2016a).

Figura 3.1 – Estrutura de um MCE



Fonte: Desenvolvido pelo autor

Observa-se no MCE da Figura 3.1, que os conceitos *A*, *B*, *C* e *D* posicionados na matriz de atributos são relacionados em causa e efeito, por meio de setas (→) que partem do conceito-causa com destino ao conceito-efeito, acrescidos de vínculos de reforço ou de balanceamento. O atributo balanceamento, simbolizado por "-", representa uma mudança entre o estado atual e o estado desejado de um conceito, gerando ações a fim de eliminar essa diferença. O atributo reforço, simbolizado por "+", indica um mecanismo que gera crescimento no sistema (ZAMBON *et al.*, 2016a).

Como exemplo de leitura do MCE da Figura 3.1, o conceito A é controlável pelo agente (*CT*), que tem dúvida se esse conceito é capaz de controlar o domínio (*PN*). Esse conceito (*A*), reduz a ação do conceito *B* por meio do balanceamento (-), cujo resultado o agente não sabe se controla (*PN*). Ocorre que esse conceito (*B*) exerce controle sobre o domínio. Por meio da leitura dessa proposição, é possível identificar uma relação indireta de poder, onde *B* é relativamente controlável pelo agente, pois ele tem controle relativo sobre *A*, que é causa de *B*. Essa relação apenas fica visível pela representação no MCE (DUARTE, 2018).

O MCE não se limita apenas a uma finalidade, podendo ser utilizado na área pública, educacional e corporativa, que demande a resolução de problemas ou o entendimento de questões complexas. A próxima seção, mostra alguns trabalhos onde os MCE foram aplicados e qual a metodologia de análise foi utilizada para compará-los.

3.2 Aplicações de MCE

O trabalho de Jaramillo (2018), focado na área de gestão pública, utilizou o MCE como um modelo de representação e comparação de modelos mentais de agentes envolvidos em iniciativa pública. O autor utiliza duas abordagens visuais para comparar MCE dos agentes. Ambas abordagens têm o objetivo de identificar conceitos em comum, a posição desses conceitos e a similaridade de suas conexões.

A primeira abordagem visa identificar mudanças no modelo mental de um mesmo agente ao longo do tempo, comparando os MCE construídos por ele. Para isso, é necessário identificar conceitos que aparecem frequentemente, o momento em que eles aparecem e desaparecem nos MCE. Quanto maior for o número de conceitos que se mantém no mapa ao longo do tempo, maior será a consistência do modelo mental.

A segunda abordagem compara MCE de diferentes agentes, visando identificar conceitos em comum que fazem parte de diferentes mapas, além da comparação do posicionamento e das conexões. Segundo o autor, a implementação de métodos de comparação entre MCE permite a identificação de modelos mentais coletivos, assim como a identificação de divergências entre modelos mentais, tanto individuais quanto coletivos.

No contexto corporativo, Zambon *et al.* (2016a) apresentou um modelo de solução de problemas em uma pequena empresa de máquinas. A empresa se encontrava em uma situação problemática no processo de entregar seus produtos aos clientes dentro do prazo combinado.

O MCE foi utilizado em um processo de duas fases interativas visando explicitar o problema da empresa, a fim de solucionar divergências entre seus dois sócios, os agentes detentores do conhecimento. A Fase I, explicitou os modelos mentais dos agentes em MCE,

por meio de entrevistas não estruturadas. Esse processo serviu para externalizar os conceitos dos agentes e permitir a identificação visual dos pontos litigantes entre eles. Na Fase II, foi possível comparar visualmente os MCE obtidos da Fase I com base nos conceitos, proposições, posicionamento e situação de controlabilidade. O processo possibilitou a negociação e a obtenção de um MCE compartilhado entre os dois agentes.

O uso do MCE, juntamente com os Estilos de Aprendizagem no trabalho de (GOMES, 2018), possibilitou acompanhar as mudanças nos Estilos de Aprendizagem dos alunos e compará-las com seu desempenho nas atividades realizadas em uma disciplina. A autora faz comparações entre MCE de alunos de um curso superior com base em relações entre conceitos no formato de *hubs*, para identificar o estilo de aprendizagem predominante de cada aluno.

O *hub* é uma estrutura significativa existente em um MCE, definido como um conceito que mantém conexões com outros conceitos, superior à média de conexões, acrescida do valor do desvio padrão, multiplicado por dois. Dentre os tipos de *hubs*, o utilizado para fazer a análise dos conceitos é o disseminador de conhecimento (HD), composto por um vértice (conceito-causa) que representa um grande emissor do fluxo.

A comparação visual utilizada nos trabalhos demanda um trabalho manual por parte do analista, requerendo um esforço e uma quantidade de tempo considerável para sua execução. Isso ocorre porque cada indivíduo tem seu modo singular de atribuir significados aos conceitos e estabelecer suas relações. Quando o analista se depara com uma grande quantidade de mapas para comparar, o problema se acentua substancialmente. A automatização do processo de comparação de MCE por meio de análises computacionais pode auxiliar o analista.

3.4 Considerações finais do capítulo

Neste capítulo foi apresentado o MCE, detalhando e descrevendo suas características e seu propósito de utilização. Diferentes trabalhos que utilizaram MCE foram apresentados para demonstrar sua relevância e flexibilidade de utilização. Observa-se que para os MCE não foram desenvolvidas técnicas computacionais que comparem suas características.

Desse modo, uma vez que um MCE é composto por um MC, o próximo capítulo discute as abordagens utilizadas por trabalhos correlatos para comparar computacionalmente MC.

4 TRABALHOS CORRELATOS

Como parte do MCE é composto por um MC, este capítulo apresenta trabalhos que propuseram algoritmos, ferramentas e funções de similaridade, para tratar a correspondência de MC de acordo com suas especificidades.

Alguns trabalhos representam a estrutura de MC como grafos para então aplicar uma abordagem de comparação (DE SOUZA *et al*., 2008; LIMONGELLI *et al.*, 2016; CHEN; LIN; CHANG, 2001; IFENTHALER, 2008). Dentre eles, existem os que mapeiam valores de atributos nos nós e arestas dos grafos, demandando a intervenção de especialistas (CHEN; LIN; CHANG, 2001) ou algoritmos de processamento de linguagem natural (DE SOUZA *et al*., 2008), para realizar o processo de comparação. Algumas abordagens utilizam MC representados como grafos acíclicos e os utilizam em suas medidas de similaridade (LIMONGELLI *et al.*, 2016; IFENTHALER, 2008). Outras necessitam que os rótulos dos conceitos sejam representados usando apenas uma letra do alfabeto (LIMONGELLI *et al.*, 2016; CHEN; LIN; CHANG, 2001) ou uma única palavra (DE SOUZA *et al*., 2008; IFENTHALER, 2008). Ainda no caso de De Souza *et al*. (2008) é necessário que os MC a serem comparados tenham o mesmo número de conceitos e termos de ligação.

Em Chen, Lin e Chang (2001) foi apresentada uma abordagem que auxiliou a comparação de MC no âmbito educacional. Os autores utilizam MC com valores pré-definidos de atributos determinados exclusivamente por especialistas dentro de um domínio específico e mapeados aos conceitos e seus relacionamentos, seguindo teorias de representação do conhecimento, aprendizagem construtiva e aprendizagem significativa. Os mapas são definidos como *Attributed Concept Maps* (ACM) e são representados como grafos dirigidos. Os conceitos são rotulados com uma letra do alfabeto junto a um valor pré-definido de atributo e correspondem aos nós do grafo. Os relacionamentos também recebem um valor de atributo, correspondendo às arestas. Os ACM são comparados a um mapa mestre (*master map*), que é construído por meio da integração de ACM desenvolvidos por especialistas e uma técnica *fuzzy*, para um determinado assunto. O valor de atributo de um conceito indica sua importância em relação aos outros conceitos do conjunto e, quanto maior seu valor, mais importante esse conceito será no conjunto. O processo de comparação utiliza a lógica estrutural proposta por

Goldsmith (1991), que aproveita as formas proposicionais e as estruturas hierárquicas dos MC com a técnica *fuzzy* proposta pelos autores. São considerados intersecção, união e vizinhança dos nós de cada mapa. O resultado gera um índice de proximidade, que indica o grau de similaridade entre dois mapas.

Em De Souza *et al*. (2008) foi utilizada correspondência de grafos para comparar dois MC. É uma abordagem que faz uso do algoritmo GIP (*Graph Isomorphism Problem*) que tem como objetivo identificar as semelhanças entre grafos de mesmo tamanho, considerando suas estruturas e atributos associados ao seus nós e linhas. Em conjunto com o GIP, o algoritmo heurístico GRASP (*Greedy Randomized Adaptive Search Procedure*) responsável por gerar boas soluções para um problema em um tempo de processamento razoável, são utilizados para a correspondência entre MC que contenham o mesmo número de conceitos e termos de ligação. Para realizar o processo de comparação dos mapas, são utilizadas duas matrizes de similaridade, uma para os conceitos e outra para os termos de ligação. Para a geração das matrizes, a similaridade tanto de conceitos quanto de termos de ligação é calculada por meio de algoritmos de processamento de linguagem natural que comparam *Strings*, como o *stemming*, desambiguação e de busca em árvore de sinônimos criada a partir do *WordNet*, resultando em um valor numérico que indica o quanto duas palavras são semanticamente similares. A abordagem permite utilizar algoritmos de comparação de grafos, heurísticos e exatos para efetuar automaticamente a identificação de similaridades entre MC.

O trabalho de Ifenthaler (2008) apresenta a tecnologia *SMD* (*Surface*, *Matching*, *Deep Structure*), um sistema que mede aspectos relacionais, estruturais e níveis semânticos para MC e representações gráficas. A tecnologia é baseada na teoria de grafos e modelos mentais, utilizando MC ou expressões de linguagem natural para analisar processos complexos. No processo computacional, os dados brutos extraídos de um MC (representado como um grafo dirigido ou não dirigido) são armazenados em pares (como uma proposição), incluindo o número do mapa ao qual a proposição pertence, o *nó1* (conceito1) como o primeiro nó da proposição, o *nó2* (conceito2) que é conectado ao *nó1* e o link que descreve a ligação entre os dois nós. Para analisar um MC, o sistema calcula três indicadores numéricos quantitativos entre todos os nós e ligações do mapa (Superfície, Correspondência e Estrutura Profunda). A Estrutura de Superfície é um indicador simples obtido pela soma de todas as proposições de um MC. A Estrutura de Correspondência é um índice calculado sobre uma árvore de extensão, contendo um subconjunto de arestas que mantém todos os nós do grafo conectados, possuindo

uma estrutura acíclica e não dirigida. O cálculo visa encontrar o diâmetro da árvore de extensão, que consiste na quantidade *links* do caminho mais curto entre os nós mais distantes da árvore. O processo se inicia transformando a árvore de extensão em uma matriz de distância D e, então, o índice da Estrutura de Correspondência é obtido pelo valor máximo de todas as entradas da matriz de distância D. A Estrutura Profunda trata da parte semântica do MC, sendo seu cálculo baseado na medida de similaridade de Tversky (1977) entre um MC e um MC de referência. O cálculo toma como base as proposições dos MC envolvidos, considerando 1, se as proposições forem iguais, e 0 caso contrário.

Em Limongelli *et al.* (2016) são propostas medidas de similaridade para a comparação de MC tratados como grafos, considerando o contexto educacional. Sete medidas de similaridade foram aplicadas e avaliadas em MC educacionais. São apresentadas maneiras diferentes e independentes de comparar dois MC. Um MC é representado por um grafo acíclico dirigido, onde seus nós, representados por uma letra, são os conceitos e as arestas são as relações entre esses conceitos. As comparações entre os mapas são feitas aplicando as medidas de similaridade em um conjunto de nós comuns, obtidos entre dois MC. As medidas não tratam o conteúdo do rótulo dos conceitos e consideram a parte estrutural dos MC para o cálculo de similaridade entre pares de nós comuns, sendo algumas híbridas, ou seja, o resultado de uma medida é utilizado em outra. Essas medidas envolvem posicionamento de nós, distância entre nós, fluxo em nós, que representa a quantidade de informação que passa por um nó e é baseada na técnica de ativação de propagação (CRESTANI, 1997), e a razão de nós predecessores, que envolve a proporção de nós predecessores comuns sobre o total de nós predecessores do grafo.

Alguns trabalhos recorrem a outros tipos de abordagens que não utilizam grafos como parte da análise de comparação de MC (CALDAS; FAVERO, 2009; SILVA *et al.*, 2011). Esses trabalhos utilizam um MC de referência (CALDAS; FAVERO, 2009) ou ontologia de referência (SILVA *et al.*, 2011) em suas análises e dependem de professores ou especialistas para concluírem o processo de comparação. O foco desses trabalhos está nas correspondências entre os rótulos de seus elementos (conceitos, termos de ligação e proposições) representados por uma palavra.

Considerando o âmbito educacional de Ensino a Distância (EAD), Caldas e Favero (2009) apresentam uma ferramenta que realiza a avaliação quantitativa e qualitativa de MC. A avaliação quantitativa utiliza a técnica de N-gramas, que calcula a similaridade de duplas de

conceito-*link* ou *link*-conceito (bi-gramas) e proposições conceito-*link*-conceito ou *link*-conceito-*link* (tri-gramas). Essa análise de similaridade é aplicada comparando as informações dos MC de estudantes em relação aos MC modelo, utilizados como referência na avaliação. Os MC modelo são formados pela comparação preliminar entre MC de alunos e MC de referência, criados por docentes ou especialistas. O resultado da análise N-gramas gera um índice de similaridade para cada mapa, e em seguida, os índices são submetidos a uma consulta do tipo *kNN* (*k-Nearest Neighbor*), que resulta em um escore para cada MC comparado. A avaliação qualitativa auxilia um estudante no processo de construção de seu MC, a fim de permitir a evolução de seu conhecimento por meio de melhorias em seu MC. Um relatório é fornecido ao estudante após a submissão de seu MC, esse relatório conta com todas as informações que não foram consideradas no MC do estudante e que estão presentes nos MC modelo. O relatório é gerado com auxílio de um dicionário de sinônimos que realiza a comparação de conceitos e sentidos das palavras de ligação. Ambos os processos de avaliação envolvem a comparação de MC desenvolvidos por alunos contra MC de referência criados por docentes ou especialistas.

A proposta de uma avaliação de aprendizagem apoiada na utilização de MC como uma ferramenta de explicitação do conhecimento de um aluno em um domínio específico e o alinhamento de ontologias para comparação automática desses mapas com uma ontologia de referência foi evidenciada no trabalho de Silva *et al.* (2011). O processo consiste na construção de um MC por um aluno acerca de um tema específico proposto por um tutor (considerando um sistema de EAD) ou professor. Em seguida, é realizado o alinhamento do mapa com uma ontologia de referência. Para que esse alinhamento seja possível, é necessário que o MC do aluno e a ontologia de referência estejam codificados na linguagem *OWL*. O MC pode ser facilmente convertido em uma ontologia com a utilização de um editor de ontologias, como o Protégé[2]. O alinhamento realizado é processado pela ferramenta SemMatcher (PIRES *et al.*, 2009), que gera correspondências entre pares de conceitos, sendo cada conceito representado como uma única palavra, combinando diferentes estratégias de correspondência (linguística e semântica).

A Tabela 4.1 apresenta uma síntese dos trabalhos apresentados levando em consideração suas características gerais. Os trabalhos foram divididos em duas categorias, os que representam a estrutura do MC como grafos e os que não representam MC como grafos.

[2] Disponível em: http://protege.stanford.edu/.

Tabela 4.1 – Síntese dos trabalhos apresentados

<table>
<tr><th colspan="2">Representam MC como Grafos</th></tr>
<tr><th>Trabalhos</th><th>Características gerais</th></tr>
<tr><td>Chen, Lin e Chang (2001) *</td><td rowspan="2">- Contexto específico (Educacional)
- Grafos acíclicos dirigidos
- Conceitos rotulados como uma letra do alfabeto
- Não trata o conteúdo do rótulo dos conceitos
* Dependente de especialistas</td></tr>
<tr><td>Limongelli et al. (2016)</td></tr>
<tr><td>De Souza et al. (2008) **</td><td rowspan="2">- Conceitos rotulados como uma única palavra
** Mesmo número de conceitos e termos de ligação para os MC comparados
** Algoritmos de processamento de linguagem natural</td></tr>
<tr><td>Ifenthaler (2008)</td></tr>
<tr><th colspan="2">Não utilizam GRAFOS para representar MC</th></tr>
<tr><th>Trabalhos</th><th>Características gerais</th></tr>
<tr><td>Caldas e Favero (2009)</td><td rowspan="2">- Contexto específico (Educacional)
- Depende de um professor ou um especialista para o processo (MC ou Ontologia de referência)
- Conceitos representados por uma palavra</td></tr>
<tr><td>Silva et al. (2011)</td></tr>
</table>

Fonte: Desenvolvido pelo autor

A abordagem proposta por este trabalho considera a correspondência de MCE de modo mais genérico em relação ao domínio de utilização dos MC, baseando-se em suas características estruturais. Considera os conceitos como sintagmas nominais, ou seja, compostos por mais de uma palavra. Os MCE podem ser hierárquicos, cíclicos ou híbridos. Para as comparações considera-se um MCE de referência, que pode ou não ser construído por um especialista no domínio. Considerando um conjunto de MCE, por meio das consultas por similaridade é possível identificar os MCE mais próximos (similares) ao MCE de referência.

4.1 Considerações finais do capítulo

Este capítulo apresentou um levantamento bibliográfico de trabalhos com metodologias para comparação de MC para auxiliarem na abordagem proposta para MCE. Diante dessas considerações, o próximo capítulo define a abordagem de análise por similaridade a ser utilizada para MCE.

5 ANÁLISE DE MCE BASEADA EM SIMILARIDADE

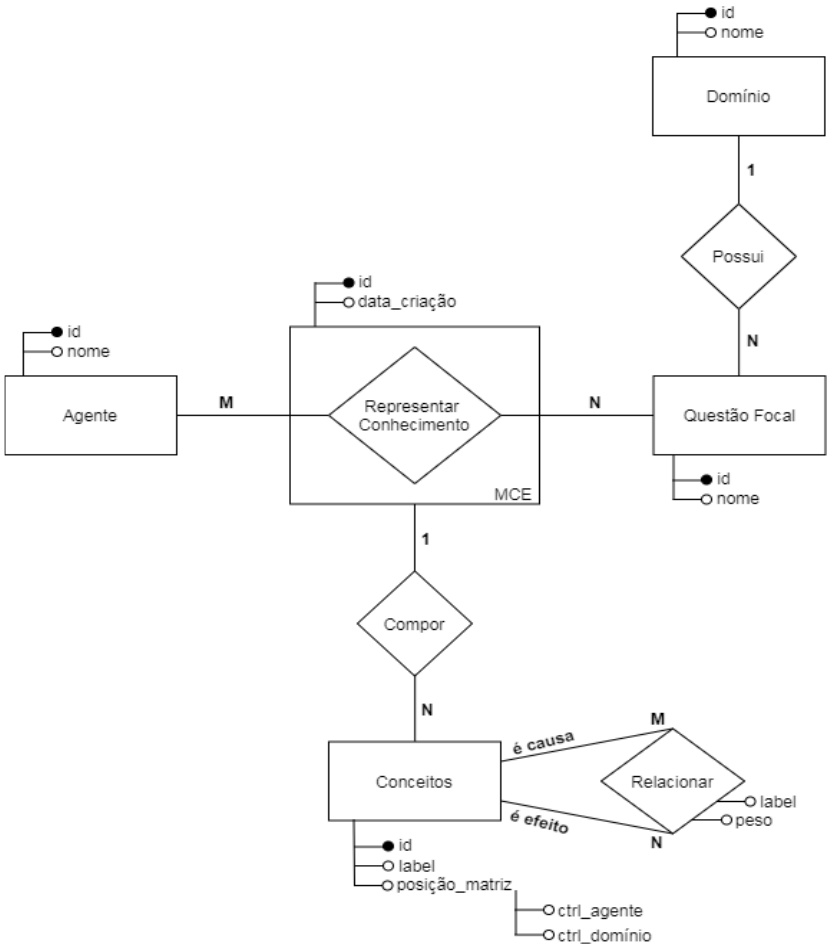
Este capítulo apresenta a abordagem para análise de MCE baseada em consultas por similaridade proposta por este trabalho. Parte-se da fundamentação teórica para a definição de MCE como um tipo de dado complexo e seu armazenamento em banco de dados (Seção 5.1). As consultas por similaridade à MCE são apresentadas na Seção 5.2. Em seguida, a Seção 5.3 apresenta as propostas de medidas de similaridade para comparação de MCE.

5.1 MCE como um tipo de dado complexo

Como abordado no Capítulo 3, um MCE se caracteriza por sua estrutura complexa, composta por um MC, eventualmente com relações de retroalimentação, posicionado em uma matriz de atributos e direcionados por pesos. Essa estrutura interna e a necessidade de armazenagem em bancos de dados para a implementação em sistemas computacionais, caracteriza um MCE como um tipo de dado complexo (Seção 2.3).

A estrutura interna do MCE apresenta partes estruturadas e não estruturadas, que podem ser observadas no esquema conceitual do banco de dados da Figura 5.1, modelado usando o MER. Na Figura 5.1, o MCE é estruturado como uma agregação que relaciona o agente e a questão focal, sendo composto por conceitos relacionados. O conjunto de entidades "Conceitos" possui os atributos: *label*, que armazena um substantivo ou sintagma nominal, e *posição_matriz*, atributo composto, para armazenar informações relativas a controlabilidade do agente e sobre o domínio. As relações de causa-efeito entre conceitos, representados pelo conjunto de relacionamentos "Relacionar", armazenam em seus atributos um verbo ou sintagma verbal – *label*, e balanceamento ou reforço – *peso*. De modo genérico, o atributo *label* presente em "Conceitos" e "Relacionar", pode ser uma palavra ou uma frase curta, o que caracteriza a parte não estruturada do MCE. Sendo assim, quanto a sua estrutura interna, o MCE pode ser classificado como um tipo de dado complexo parcialmente estruturado e não estruturado.

Figura 5.1 – Esquema conceitual do banco de dados para armazenamento de MCE



Fonte: Desenvolvido pelo autor

Ainda na Figura 5.1, pode-se observar, pela cardinalidade do conjunto de relacionamentos “Compor” (1:N), que um MCE pode ser composto por N conceitos, relacionados entre si com cardinalidade M:N (conjunto de relacionamentos “Relacionar”). Portanto, cada MCE armazenado poderá ter número variável de conceitos e relacionamentos. Desse modo, quanto ao domínio, o MCE pode ser classificado como um tipo de dado complexo adimensional.

Considerando a estrutura interna de MCE é possível observar que dificilmente, ou quase nunca, dois MCE de dois agentes distintos serão iguais. Desse modo, critérios de igualdade e ordem não são adequados para a comparação de MCE. Como MCE são tipos de dados complexos, então pode-se utilizar o critério de similaridade nas comparações (Seção 2.3).

Assim, a partir do armazenamento de MCE em banco de dados, suas características podem ser utilizadas nas operações de comparação efetuadas para responder consultas por similaridade.

5.2 Consultas por similaridade aos MCE

Uma vez que MCE são dados complexos devidamente armazenados em um banco de dados, é possível consultá-los por similaridade para diversas aplicações. Com isso, pode-se considerar duas categorias principais de análise, que são a base para possíveis aplicações do MCE:

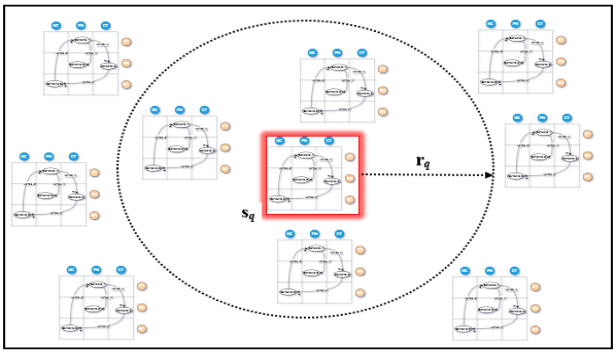
1. Análise de aprendizado: esta categoria pode avaliar o aprendizado de pessoas em diversos âmbitos. Em um exemplo prático de aplicação, um professor pode construir um MCE sobre um determinado assunto e pedir para que seus alunos façam o mesmo. A partir da comparação dos MCE dos alunos com o do professor, será verificado quais são os mais similares, podendo ser considerados como parte da avaliação a que os alunos são submetidos.

2. Análise de padrões: esta categoria pode detectar ou buscar padrões específicos para diversas necessidades. Por exemplo, uma empresa busca suprir as necessidades de um setor quanto a determinadas competências, que são descritas em um MCE. O mapa construído pode ser comparado aos MCE construídos e armazenados por candidatos sobre suas competências e, dentre eles, pode-se identificar padrões que denotem as competências desejadas.

Em ambas categorias de análise pode-se considerar um MCE de referência a ser comparado com outros MCE. Diante dessa necessidade, consultas por similaridade podem ser utilizadas (Seção 2.3).

As figuras 5.2 e 5.3 ilustram uma consulta por abrangência e uma consulta aos *k*-vizinhos mais próximos, respectivamente. Em ambas, considera-se que o objeto central de busca s_q é o MCE de referência, a ser comparado por similaridade com todos os outros MCE armazenados. Desse modo, a consulta por abrangência ilustrada na Figura 5.2, retornará os MCE cuja similaridade ao MCE de referência (s_q) esteja dentro do raio de busca r_q.

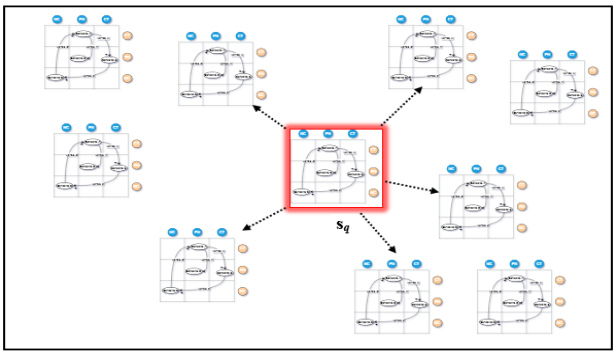
Figura 5.2 – Exemplo de consulta por abrangência



Fonte: Desenvolvido pelo autor

Por sua vez, a consulta aos *k*-vizinhos mais próximos, ilustrada na Figura 5.3, retornará os cinco MCE mais próximos ($k = 5$) ao MCE base (s_q).

Figura 5.3 – Exemplo de consulta aos k-vizinhos mais próximos



Fonte: Desenvolvido pelo autor

Para as consultas por similaridade é necessário comparar MCE, como abordado na próxima seção.

5.3 Comparação por similaridade de MCE

Para realizar as comparações por similaridade de MCE, propõe-se utilizar técnicas de correspondência complexas, relacionando as categorias das técnicas de correspondência descritas na Seção 2.2 com as características do MCE. O objetivo é propor funções de similaridade que possam considerar parte ou todas as características dos MCE, de modo que o analista possa escolher qual utilizar.

Como abordado anteriormente, o MCE é um tipo de dado complexo parte estruturado e parte não estruturado, do qual é possível extrair os seguintes vetores de características (*VC*) adimensionais:

VC1 = [conceitos com posicionamentos];

VC2 = [relacionamentos com pesos];

VC3 = [proposições]

Considerando dois MCE a serem comparados (*MCE1* e *MCE2*) e os três *VC*, são propostas as seguintes possibilidades para a função de similaridade (*Fs*):

$Fs(VC1_{MCE1}, VC1_{MCE2})$ – retorna o grau de similaridade entre *MCE1* e *MCE2* a partir de seus *VC1*, ou seja, considera apenas o grau de similaridade dos rótulos dos conceitos presentes nos MCE comparados. O uso do atributo de posicionamento dos conceitos pode ser opcional;

$Fs(VC2_{MCE1}, VC2_{MCE2})$ – retorna o grau de similaridade entre *MCE1* e *MCE2* a partir de seus *VC2*, ou seja, considera o grau de similaridade de conceitos com rótulos similares e que se relacionam em causa e efeito (proposições sem o termo de ligação). O uso do atributo de peso dos relacionamentos pode ser opcional;

$Fs(VC3_{MCE1}, VC3_{MCE2})$ – retorna o grau de similaridade entre *MCE1* e *MCE2* a partir de seus *VC3*, ou seja, considera o grau de similaridade das proposições dos MCE comparados (conceitos similares que se relacionam em causa e efeito, incluindo os rótulos dos termos de ligação). O uso do atributo de peso dos relacionamentos pode ser opcional.

Como abordado na Seção 2.2, é possível construir técnicas de correspondência complexas a partir da combinação de técnicas de correspondência de diferentes categorias, de modo a se obter um resultado eficaz. As possíveis *Fs* podem ser obtidas a partir de três categorias de correspondência: *Graph-based*, *String-based* e *Language-based*.

A categoria *Graph-based* possibilita a comparação da parte estruturada do MCE, sendo possível estabelecer uma equivalência com a estrutura de um grafo, onde os vértices correspondem aos conceitos (*VC1*) e as arestas aos relacionamentos (*VC2*) entre conceitos. Agregando aos conceitos os seus atributos de posicionamento na matriz e aos relacionamentos os seus pesos de balanceamento ou reforço, propõe-se as três *Fs*. As categorias *String-based* e *Language-based* contemplam a comparação da parte não estruturada do MCE, que corresponde aos rótulos dos conceitos (*VC1*) e dos termos de ligação das proposições (*VC3*). Em ambos os casos existe a necessidade de comparar *Strings*. A categoria *String-based* possibilita a comparação por similaridade sintática dos rótulos dos conceitos e dos termos de ligação, usando algoritmos como *Jaro*, *TFIDF*, *Leveshtein etc*. Por sua vez, a categoria *Language-based* apresenta técnicas baseadas no processamento de linguagem natural, que possibilitam a comparação por similaridade considerando a semântica dos rótulos dos conceitos e dos termos de ligação. Assim, é possível propor $Fs(VC1_{MCE1}, VC1_{MCE2})$, por meio de técnicas das categorias *String-based* e/ou *Language-based*, sendo que $Fs(VC2_{MCE1}, VC2_{MCE2})$, $Fs(VC3_{MCE1}, VC3_{MCE2})$, necessitam também de técnicas da categoria *Graph-based*. Para uma comparação por similaridade que considera todas as características dos MCE, o analista poderá utilizar nas consultas $Fs(VC3_{MCE1}, VC3_{MCE2})$. Dependendo da aplicação, o analista poderá considerar menos características dos MCE para realizar sua análise. Por exemplo, o uso de $Fs(VC1_{MCE1}, VC1_{MCE2})$ possibilita uma abordagem de comparação que considera somente os rótulos de conceitos do MCE. De modo geral, quanto maior o número de características consideradas nas comparações de MCE, maior será a acurácia do resultado, e as consultas retornarão um conjunto menor de MCE similares. Entretanto, maior acurácia implica em menor eficiência em termos de tempo de execução dos algoritmos das *Fs*. Por exemplo, técnicas da categoria *Language-based* necessitam de pré-processamento das *Strings* a serem comparadas e consultas a recursos externos (como dicionário de sinônimos), o que tende a aumentar o tempo de execução. Por outro lado, a eficiência também depende do número de MCE a serem comparados nas consultas por similaridade.

5.4 Considerações finais do capítulo

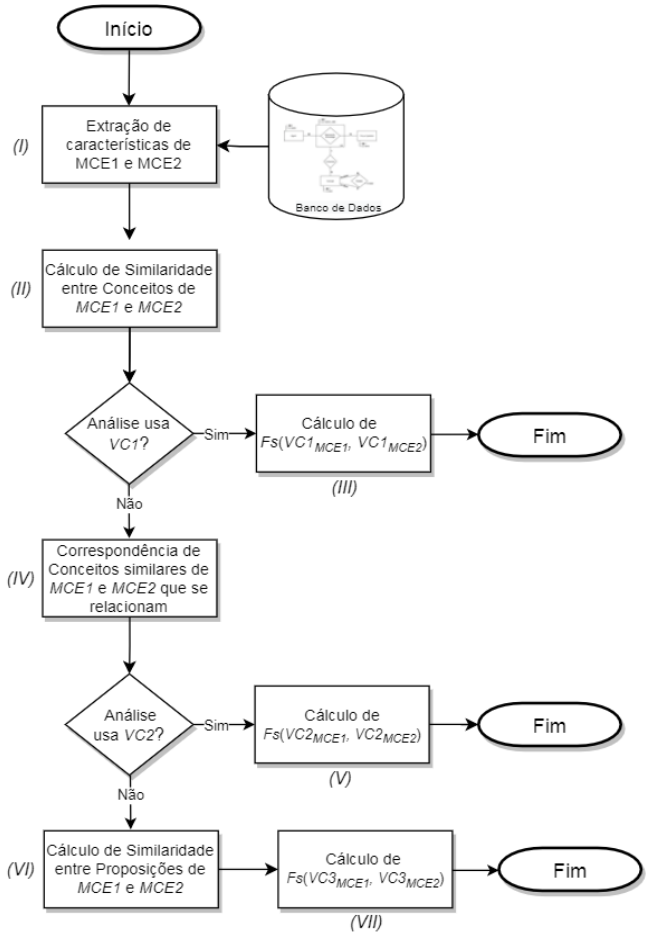
A proposta apresentada permite ao analista, a partir da escolha das características a serem utilizadas pelas funções de similaridade e do número de MC a serem analisados, determinar o quanto de eficácia considera ser necessária no resultado das consultas por

similaridade, ponderando acurácia e eficiência. O próximo capítulo apresenta o processo de comparação de MCE utilizando a abordagem apresentada.

6 PROCESSO DE COMPARAÇÃO DE MCE

De acordo com a abordagem para análise de MCE proposta no Capítulo 5, este capítulo apresenta o processo de comparação por similaridade de MCE, retornando o grau de similaridade dos MCE comparados para qualquer *Fs* selecionada por um analista (Seção 5.2). As *Fs* definidas no processo são aplicáveis aos algoritmos de consulta por similaridade *RQ* e *kNNQ* (Seção 5.2). O processo de comparação por similaridade de MCE é composto por sete etapas, como mostra a Figura 6.1 *(I) Extração de Características de MCE1 e MCE2*, *(II) Cálculo de Similaridade entre Conceitos de MCE1 e MCE2*, *(III) Cálculo de* $Fs(VC1_{MCE1}, VC1_{MCE2})$, *(IV) Correspondência de Conceitos similares de MCE1 e MCE2 que se relacionam, (V) Cálculo de* $Fs(VC2_{MCE1}, VC2_{MCE2})$, *(VI) Cálculo de Similaridade entre Proposições de MCE1* e *MCE2*, *(VII) Cálculo de* $Fs(VC3_{MCE1}, VC3_{MCE2})$. Essas etapas são abordadas a seguir.

Figura 6.1 – Fluxograma do processo de comparação de MCE



Fonte: Desenvolvido pelo autor

(I) Extração de Características de MCE1 e MCE2

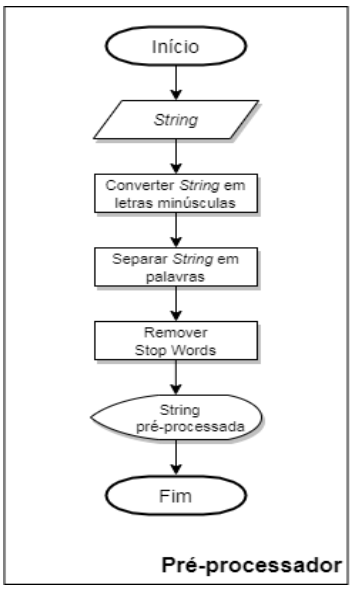
A partir dos MCE armazenados em um banco de dados de acordo com o esquema conceitual proposto no Capítulo 5 (Figura 5.1), seleciona-se os dois MCE (*MCE1* e *MCE2*) a serem comparados por similaridade, extraindo suas características (conceitos com posicionamentos, relacionamentos com pesos e proposições). Desse modo, esta etapa resulta nos vetores de características *VC1*, *VC2* e *VC3* de *MCE1* e *MCE2* a serem utilizados nas próximas etapas do processo.

(II) Cálculo de Similaridade entre Conceitos de MCE1 e MCE2

Os rótulos dos conceitos (*VC1*) e dos termos de ligação das proposições (*VC3*), que correspondem à parte não estruturada do MCE, são comparados por meio de técnicas de correspondência da categoria *Language-based* (Capítulo 2, Seção 2.2).

Desse modo, cada conceito e termo de ligação de *MCE1* e *MCE2* necessita ser submetido a uma série de etapas de pré-processamento para se tornarem aptos a serem comparados por similaridade. Essas etapas são realizadas pelo Pré-processador, de acordo com a Figura 6.2, onde rótulos de conceitos ou de termos de ligação são tratados genericamente como *Strings*. Ressalta-se que, tanto os conceitos quanto os termos de ligação não se limitam a uma palavra, sendo considerados sintagmas nominais e verbais, respectivamente.

Figura 6.2 – Fluxograma de etapas do Pré-processador



Fonte: Desenvolvido pelo autor

De acordo com a Figura 6.2, a primeira etapa do pré-processamento converte a *String* (rótulo de conceito ou de termo de ligação) em letras minúsculas. A etapa seguinte segmenta a *String* para identificar as palavras que a compõe e, após isso, é realizada a remoção de *stop words*. Ressalta-se que a eliminação de termos irrelevantes para a comparação reduz a dimensão do problema, implicando em um ganho de eficiência computacional. Após realizado o pré-processamento, parte-se para o cálculo de similaridade de pares de *Strings* (correspondentes aos rótulos de conceitos ou de termos de ligação) por meio do Algoritmo 1.

Algoritmo 1: Algoritmo Jaccard Adaptado

Entrada: $p_x = \{p_{x1}, \ldots, p_{xm}\}$: m palavras de s_1
$p_y = \{p_{y1}, \ldots, p_{yn}\}$: n palavras de s_2

Saída: $sim_{string}(s_1, s_2)$: valor de similaridade entre [0,1]

1. para cada $p_{xi} \in s_1$ **faça**
2. **para** cada $p_{yj} \in s_2$ **faça**
3. **se** $p_{xi} == p_{yj}$ **então** $nI += 1$;
4. **senão se** p_{xi} é sinônima de p_{yj} **então** $nS += 1$;
5. **senão se** radical de p_{xi} == radical de p_{yj} **então** $nR += 1$;
6. **senão** p_{xi} e $p_{yj} = 0$;
7. **fimpara**
8. fimpara
9. calcular $sim_{string}(s_1, s_2)$ usando

$$sim_{string}(s_1, s_2) = \frac{|nI(p_x,p_y) \times \alpha + nS(p_x,p_y) \times \beta + nR(p_x,p_y) \times \delta|}{|p_x \cup p_y|}$$

10. $R \leftarrow sim_{string}$

O Algoritmo 1 calcula a similaridade entre duas *Strings* por meio da função de similaridade de Jaccard (MOREIRA *et al.*, 2015), mas com adaptações, sendo denominado Algoritmo Jaccard Adaptado. Cada palavra contida nas *Strings* s_1 e s_2 são comparadas, determinando: $nI(p_x,p_y)$ que corresponde ao número de palavras idênticas contidas nos conjuntos de palavras p_x e p_y; $nS(p_x,p_y)$ é o número de palavras sinônimas contidas em p_x e p_y; e $nR(p_x,p_y)$ é o número de palavras com o mesmo radical (processo de *stemming*) contidas em p_x e p_y. As variáveis α, β e δ são parâmetros de pesos entre [0, 1] a serem definidos pelo analista, que pondera a importância de *nI*, *nS* e *nR*. A similaridade de s_1 e s_2 é calculada pela razão entre a soma dos valores de similaridade ponderados, de palavras idênticas, sinônimas e com o mesmo radical e a união das palavras contidas em p_x e p_y. As comparações resultam em valores de similaridade entre 0 e 1, onde 1 significa que as *Strings* comparadas são totalmente iguais e 0 totalmente diferentes.

Considerando o processo de comparação de MCE apresentado na Figura 6, o cálculo de similaridade de termos de ligação será necessário na etapa *(VI)*, a ser abordada posteriormente. Na etapa *(II)* é necessário calcular a similaridade entre os conceitos de *MCE1* e *MCE2*. Para isso, cada rótulo de conceito de *MC1* é comparado com todos os rótulos de conceitos de *MCE2* usando o Algoritmo 1. Os pares de conceitos resultantes das comparações, com os respectivos valores de similaridade, são armazenados em um vetor *R*, a ser utilizado como entrada do Algoritmo 2.

O Algoritmo 2, denominado de Algoritmo Alinhamento C, associa cada rótulo de conceito de *MCE1* a apenas um rótulo de conceito de *MCE2* (1:1), considerando o maior valor de similaridade entre eles. O algoritmo também verifica o posicionamento dos conceitos na matriz de atributos, considerando sua controlabilidade do agente e sobre o domínio. Essa verificação só ocorre se a variável *posC* receber o valor 1, o que significa que o atributo de posicionamento dos conceitos está sendo considerado no algoritmo. Desse modo, a saída do algoritmo é o vetor R_{VC1} que contém as melhores correspondências (maior valor de similaridade) entre os conceitos de *MCE1* e *MCE2* e, no caso de *posC* = 1, o valor de similaridade obtido já considera os posicionamentos também. O vetor R_{VC1} é necessário para as etapas *(III)* e *(IV)* do processo de comparação de MCE (Figura 6).

O vetor R_{VC1} é obtido a partir dos resultados de similaridade dos pares de conceitos armazenados no vetor *R* (entrada do Algoritmo 2). Primeiramente, o processo busca em *R* pares de conceitos com valores de similaridade 1 (conceitos idênticos em *MCE1* e *MCE2*), inserindo-os em R_{VC1} e, então, remove de *R* todas as ocorrências desses conceitos. Em seguida, busca-se em *R* os pares de conceitos com maiores valores de similaridade a serem inseridos em R_{VC1}, ou seja, para cada conceito de *MCE1*, busca-se o maior valor de similaridade deste conceito em relação aos conceitos de *MCE2*. Em caso de empate, ou seja, o conceito de *MCE1* possui o mesmo grau de similaridade com mais de um conceito de *MCE2*, considera-se a primeira ocorrência. No caso de sobrarem conceitos em um dos mapas ou o maior valor de similaridade de todas as comparações for igual 0, os conceitos serão associados ao termo "*sem correspondência"* ao serem inseridos em R_{VC1}.

Algoritmo 2: Algoritmo Alinhamento C

Entrada:
$R = \{(c_{xi}, c_{yj}, sim), \ldots, (c_{xm}, c_{yn}, sim)\}$: n pares de conceitos com valores de similaridade
var posC : boolean
Saída:
$R_{VC1} = \{(c_{xi}, c_{yj}, sim), \ldots, (c_{xm}, c_{yn}, sim)\}$: vetor com melhores correspondências de conceitos (1:1)

1. para cada $c_{xi} \in R$ **faça**
2. **para** cada $c_{yj} \in R$ **faça**
3. **se** $sim(c_{xi}, c_{yj}) = 1$ **então** inserir em R_{VC1}
4. obter o maior valor de similaridade para (c_{xi}, c_{yj})
5. $R_{VC1} \leftarrow (c_{xi}, c_{yj}, sim)$
6. **se** c_{xi} possuir o mesmo grau de similaridade com mais de um c_{yj} **então** considerar a primeira ocorrência de c_{xi}
7. **fimpara**
8. fimpara
9. se restarem c_{xi} ou c_{yj} em algum dos mapas || maior valor de similaridade de $(c_{xi}, c_{yj}) = 0$ **então**
$R_{VC1} \leftarrow (c_{xi}$ ou c_{yj}, *"sem correspondência"*)
10. se posC == 1 **então**
11. para cada $(c_{xi}, c_{yj}) \in R_{VC1}$ **faça**
12. obter $sim(c_{xi}, c_{yj})$ de R_{VC1}
13. obter os posicionamentos dos conceitos pos_{cxi} e pos_{cyj} de R_{VC1}
14. **se** controlabilidade do agente de pos_{cxi} e pos_{cyj} são iguais && controlabilidade do domínio de pos_{cxi} e pos_{cyj} são iguais **então** $sim(pos_{cxi}, pos_{cyj}) = 1$
15. **se** controlabilidade do agente de pos_{cxi} e pos_{cyj} são iguais && controlabilidade do domínio de pos_{cxi} e pos_{cyj} são diferentes || **se** controlabilidade do agente de pos_{cxi} e pos_{cyj} são diferentes && controlabilidade do domínio de pos_{cxi} e pos_{cyj} são iguais **então** $sim(pos_{cxi}, pos_{cyj}) = 0.5$
16. **senão** $sim(pos_{cxi}, pos_{cyj}) = 0$
17. calcular sim_{con} usando

$$\boldsymbol{sim_{con} = \frac{sim(c_{xi}, c_{yj}) + sim(pos_{cxi}, pos_{cyj})}{2}}$$

18. $\boldsymbol{R_{VC1} \leftarrow sim_{con}}$
19. fimpara

Ainda no Algoritmo 2, depois de obter os melhores valores de similaridade para os conceitos, busca-se em R_{VC1} o posicionamento de cada um, verificando seus atributos de posicionamento em relação ao agente e ao domínio. O valor de similaridade para os posicionamentos é obtido verificando se os conceitos possuem os mesmos atributos de posicionamento em relação ao agente e ao domínio. Se dois atributos de posicionamento forem iguais, significa que os conceitos estão alocados no mesmo quadrante da matriz de atributos, e então, o valor de similaridade é 1. Caso apenas um atributo de posicionamento entre os conceitos for igual, o valor de similaridade entre eles é 0.5. Por fim, se não apresentarem

nenhuma igualdade de posicionamento, o valor de similaridade é 0. Então, a similaridade é calculada pela média aritmética do valor de similaridade dos rótulos dos conceitos presente em R_{VC1} e o resultado da similaridade do posicionamento desses conceitos na matriz de atributos. O valor de similaridade armazenado em R_{VC1} é, então, atualizado. Ressalta-se que, caso o analista não deseje considerar o posicionamento dos conceitos no cálculo da similaridade (*posC* = 0), então o vetor R_{VC1} não sofrerá alterações, mantendo-se o valor de similaridade obtido apenas considerando os conceitos.

(III) Cálculo de $Fs(VC1_{MCE1}, VC1_{MCE2})$

Considerando os rótulos dos conceitos (*VC1)* e seus posicionamentos para o cálculo de similaridade entre *MCE1* e *MCE2*, a $Fs(VC1_{\mathrm{MCE1}}, VC1_{\mathrm{MCE2}})$ é obtida em (1). O cálculo envolve razão entre a soma de todos os valores de similaridade dos pares de conceitos contidos em R_{VC1} pela quantidade total de pares de conceitos (N_{con}).

$$Fs(VC1_{MCE1}, VC1_{MCE2}) = \frac{\sum_{i=1}^{n} sim_{coni}}{N_{con}} \quad (1)$$

Lembrando que *VC1* contém os conceitos com os posicionamentos, sendo que o uso ou não dos atributos de posicionamentos foi definido na etapa *(II)*. O analista pode considerar mais características no processo de comparação de MCE (Figura 6). Desse modo, o processo é continuado na etapa *(IV)*.

(IV) Correspondência de Conceitos similares de MCE1 e MCE2 que se relacionam

Até a etapa *(III)* do processo de comparação de MCE (Figura 6) utilizou-se apenas os conceitos (*VC1*), que correspondem a parte não estruturada dos mapas. A partir da etapa *(IV)* considera-se também a parte estruturada dos MCE, que é equivalente a estrutura de um grafo. Desse modo, utiliza-se a combinação de técnicas de correspondência das categorias *Language-based* e *Graph-based* (Capítulo 2, Seção 2.2).

A etapa *(IV)* é responsável por encontrar relacionamentos similares entre *MCE1* e *MCE2*, ou seja, conceitos com rótulos similares e que se relacionam em causa e efeito (*VC2*). Para isso é necessário comparar por similaridade os conceitos-causa e os conceitos-efeito de cada relacionamento de *MCE1* com os conceitos-causa e os conceitos-efeito de todos os

relacionamentos de *MCE2*. Dois relacionamentos serão considerados similares quando seus conceitos-causa e conceitos-efeito forem similares, ou seja, tiverem valor de similaridade maior que zero. A similaridade entre dois relacionamentos é, então, calculada pela média aritmética do valor da similaridade entre seus conceitos-causa e seus conceitos-efeito. Esse processo é realizado pelo Algoritmo 3, denominado Algoritmo Alinhamento Rel.

Algoritmo 3: Algoritmo Alinhamento Rel

Entrada:
$R_{VC1} = \{(c_{xi}, c_{yj}, sim), \ldots, (c_{xn}, c_{yn}, sim)\}$: vetor com melhores correspondências de conceitos (1:1)
$VC2_{MCE1} = \{(r_{xi}, peso_{rxi}), \ldots, (r_{xn}, peso_{rxn})\}$: vetor relacionamentos com pesos de *MCE1*
$VC2_{MCE2} = \{(r_{yj}, peso_{ryj}), \ldots, (r_{yn}, peso_{yn})\}$: vetor relacionamentos com pesos de *MCE2*
var relC : boolean
Saída:
$R_{VC2} = \{(r_{xi}, r_{yj}, sim), \ldots, (r_{xn}, r_{yn}, sim)\}$: vetor com correspondências de relacionamentos

1. para cada $r_{xi} \in VC2_{MCE1}$ **faça**
2. **para** cada $r_{yj} \in VC2_{MCE2}$ **faça**
3. **se** *sim* de conceitos-causa (c_{xi}, c_{yi}) de $r_{xi}, r_{yj} > 0$ && *sim* de conceitos-efeito (c_{xj}, c_{yj}) de $r_{xi}, r_{yj} > 0$ em R_{VC1} **então** calcular sim_{rel} usando

$$sim_{rel} = \frac{sim(c_{xi}, c_{yi}) + sim(c_{xj}, c_{yj})}{2}$$

4. $R_{VC2} \leftarrow sim_{rel}$
5. **fimpara**
6. fimpara
7. se relC == 1 **então**
8. **para** cada $(r_{xi}, r_{yj}) \in R_{VC2}$ **faça**
9. **se** *peso* de r_{xi} é igual a *peso* de r_{yj} **então** $sim(peso_{rxi}, peso_{ryj}) = 1$
10. **senão** $sim(peso_{rxi}, peso_{ryj}) = 0$
11. calcular sim_{relp} usando

$$sim_{rel} = \frac{sim(c_{xi}, c_{yi}) + sim(c_{xj}, c_{yj}) + sim(peso_{rxi}, peso_{ryj})}{3}$$

12. $R_{VC2} \leftarrow sim_{rel}$
13. **fimpara**

Ainda no Algoritmo 3, é verificado se os relacionamentos obtidos em R_{VC2} possuem o mesmo peso de balanceamento ou reforço. Então, a similaridade entre relacionamentos com pesos é calculada pela média aritmética de cada valor de similaridade presente em R_{VC2} e os pesos associados a cada relacionamento. O algoritmo recebe como entrada os vetores $VC2_{MCE1}$ e $VC2_{MCE2}$ que contém relacionamentos com pesos de *MCE1* e *MCE2*, respectivamente, e a

variável *relC* que, quando receber o valor 1, o algoritmo considera os pesos dos relacionamentos. Ou seja, caso o analista não deseje considerar os pesos dos relacionamentos no cálculo da similaridade, basta que *relC* receba o valor 0. O valor de similaridade entre conceitos-causa e conceitos-efeito presentes nos relacionamentos é obtido no vetor R_{VC1}, resultante da etapa *(II).* Os pesos de cada relacionamento são obtidos nos vetores $VC2_{MCE1}$ e $VC2_{MCE2}$. A similaridade entre os pesos é obtida por meio de uma verificação simples, se os relacionamentos comparados possuírem o mesmo peso de balanceamento ou reforço, então o valor de similaridade entre os pesos é 1, caso contrário, sua similaridade é 0. A saída do Algoritmo 3 é o vetor R_{VC2}, que contém os pares de relacionamentos com pesos e seus respectivos valores de similaridade. Apenas os pares de relacionamentos que tiverem valor de similaridade maior que zero serão armazenados em R_{VC2}. Os resultados armazenados em R_{VC2} serão utilizados nas etapas *(V)* e *(VI).*

(V) Cálculo de $Fs(VC2_{MCE1}, VC2_{MCE2})$

A etapa (*V*) calcula $Fs(VC2_{MCE1}, VC2_{MCE2})$ que retorna o grau de similaridade entre *MCE1* e *MCE2* considerando seus relacionamentos com pesos (*VC2*). De acordo com (2), o resultado é obtido por meio da razão entre a soma de todos os valores de similaridade dos relacionamentos com pesos contidos em R_{VC2} e a maior quantidade de relacionamentos (N_{rel}) entre *MCE1* e *MCE2*.

$$Fs(VC2_{MCE1}, VC2_{MCE2}) = \frac{\sum_{i=1}^{n} sim_{rel}}{N_{rel}} \quad (2)$$

Lembrando que *VC2* contém os relacionamentos com os pesos, sendo que o uso ou não dos atributos de pesos foi definido na etapa *(IV).*

(VI) Cálculo de Similaridade entre Proposições de MCE1 e MCE2

A etapa (*VI*) calcula a similaridade entre as proposições dos MCE (*VC3*). Para isso, é necessário considerar a similaridade entre os relacionamentos de *MCE1* e *MCE2*, realizado na etapa *(IV)*, incluindo os rótulos dos termos de ligação.

O cálculo de similaridade dos rótulos de termos de ligação é realizado por meio do Algoritmo 1, apresentado na etapa *(II)*, e os valores de similaridade entre relacionamentos com seus

respectivos pesos são obtidos em R_{VC2}, resultante da etapa *(VI)*. O cálculo de similaridade considera a média aritmética entre o valor de similaridade de cada relacionamento e de cada termo de ligação resultando na similaridade de cada proposição. O processo é realizado por meio do Algoritmo 4, denominado Algoritmo Alinhamento P.

Algoritmo 4: Algoritmo Alinhamento P

Entrada:
$R_{VC2} = \{(r_{xi}, r_{yj}, sim), \ldots, (r_{xn}, r_{yn}, sim)\}$: vetor com correspondências de relacionamentos
$VC3_{MCE1} = \{(p_{xi}), \ldots, (p_{xn})\}$: vetor com proposições de *MCE1*
$VC3_{MCE2} = \{(p_{yj}), \ldots, (p_{yn})\}$: vetor com proposições de *MCE2*
Saída:
$R_{VC3} = \{(p_{xi}, p_{yj}, sim), \ldots, (p_{xn}, p_{yn}, sim)\}$: vetor com correspondências de proposições

1. para cada $(r_{xi}, r_{yj}) \in R_{VC2}$ **faça**
2. obter $sim(r_{xi}, r_{yj})$ de R_{VC2}
3. obter os rótulos dos termos de ligação tl_{xi} de $VC3_{MCE1}$ e tl_{yj} de $VC3_{MCE2}$
4. calcular $sim(tl_{xi}, tl_{yj})$ utilizando o Algoritmo 1
5. calcular sim_{prop} usando

$$\boldsymbol{sim_{prop} = \frac{sim(r_{xi}, r_{yj}) + sim(tl_{xi}, tl_{yj})}{2}}$$

6. $\boldsymbol{R_{VC3} \leftarrow sim_{prop}}$

O Algoritmo 4 recebe como entrada os vetores $VC3_{MCE1}$ e $VC3_{MCE2}$ com as proposições de *MCE1* e *MCE2*, respectivamente, bem como R_{VC2}. O algoritmo resulta no vetor R_{VC3}, que contém os pares de proposições com pesos e os respectivos valores de similaridade. Os resultados armazenados em R_{VC3} serão utilizados na etapa *(VII)*, para o cálculo da última função de similaridade.

(VII) Cálculo de $Fs(VC3_{MCE1}, VC3_{MCE2})$

A penúltima etapa do processo de comparação de MCE (Figura 6) calcula $Fs(VC3_{MCE1}, VC3_{MCE2})$ que retorna o grau de similaridade entre *MCE1* e *MCE2* considerando suas proposições (*VC3*). De acordo com (3), o resultado é obtido por meio da razão entre a soma de todos os valores de similaridade das proposições contidos em R_{VC3} e a maior quantidade de proposições (N_{prop}) entre *MCE1* e *MCE2*.

$$Fs(VC3_{MCE1}, VC3_{MCE2}) = \frac{\sum_{i=1}^{n} sim_{propi}}{N_{prop}} \quad (3)$$

6.1 Considerações finais do capítulo

Este capítulo apresentou um processo para o cálculo de similaridade entre dois MCE, possibilitando ao analista considerar uma ou até todas as suas características. Sabendo que o MC está presente na estrutura do MCE, usando as *Fs* sem considerar atributos opcionais do MCE, ou seja, atribuindo 0 para as variáveis *posC* e *relC* dos algoritmos 2 e 3, é possível comparar MC, considerando uma ou até todas suas características. Desse modo, foram definidas três *Fs* as quais podem ser utilizadas em algoritmos de consultas por similaridade *RQ* e *KNNQ*, apresentados no capítulo 5 (Seção 5.2). O próximo capítulo apresenta um exemplo de aplicação do processo proposto.

7 ESTUDO DE CASO

Para ilustrar o processo de análise de MCE baseada em comparação por similaridade proposto por esta dissertação, foi realizado um experimento real no domínio educacional. Foram considerados MCE de quatro alunos e de um professor, previamente armazenados em um banco de dados de acordo com o esquema conceitual da Figura 5.1 (Capítulo 5).

Como os MCE utilizados foram descritos em língua portuguesa, o Algoritmo de Jaccard Adaptado presente na etapa *(II)* do Capítulo 5, utilizou uma API denominada JWN-Br (OLIVEIRA; ROMAN, 2013). Essa API fornece acesso ao Wordnet.Br, uma versão do *WordNet* desenvolvida para o português do Brasil (DA SILVA, 2006) que contém mais de vinte mil grupos de sinônimos, denominados *SynSets*. Duas palavras são consideradas sinônimas se uma palavra estiver contida no grupo de *SynSets* da outra. Para verificar o radical das palavras foi utilizado o algoritmo de *Stemming* proposto em (ORENGO; HUYCK, 2001) com adições de novas regras e exceções de (COELHO, 2007).

O experimento é análogo ao processo de análise de aprendizado do Capítulo 5 (Seção 5.2), onde o professor e seus alunos construíram seus MCE com base na seguinte questão focal: "*Qual a relação entre o significado de valor para a empresa e para o cliente*?"

As quantidades de conceitos e proposições dos MCE utilizados são descritas na Tabela 7.1.

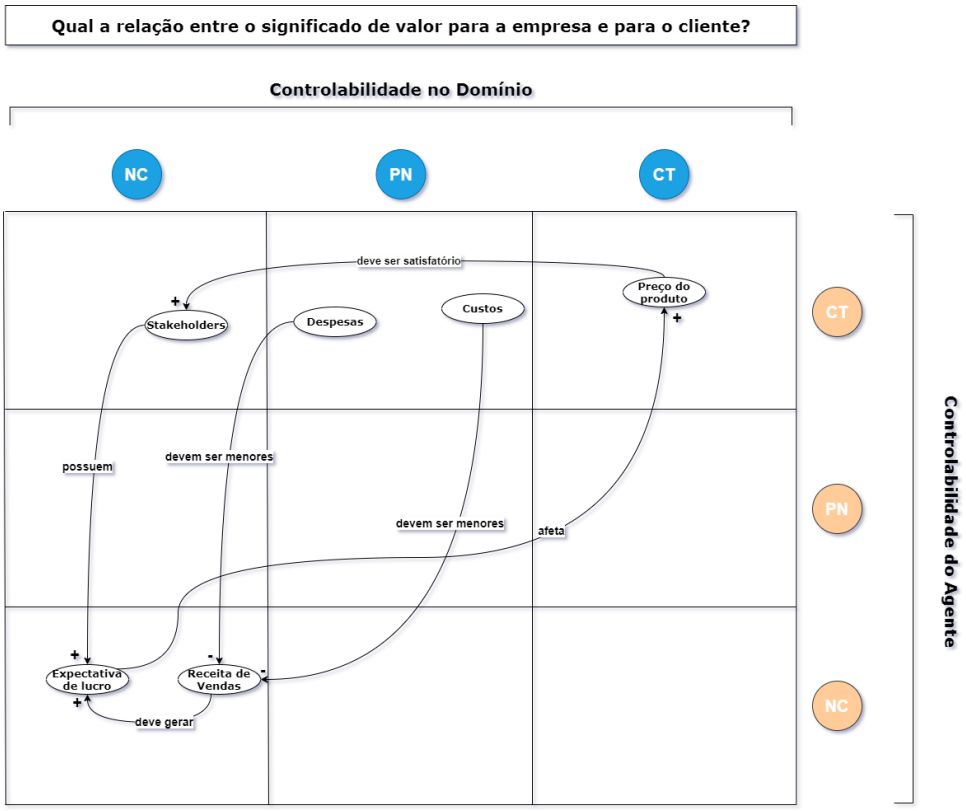
Tabela 7.1 – Quantidade de Conceitos e Proposições dos MCE

MCE	Quantidade de Conceitos	Quantidade de Proposições
Professor	17	23
Aluno 1	6	6
Aluno 2	7	7
Aluno 3	11	10
Aluno 4	8	7

Fonte: Desenvolvido pelo autor

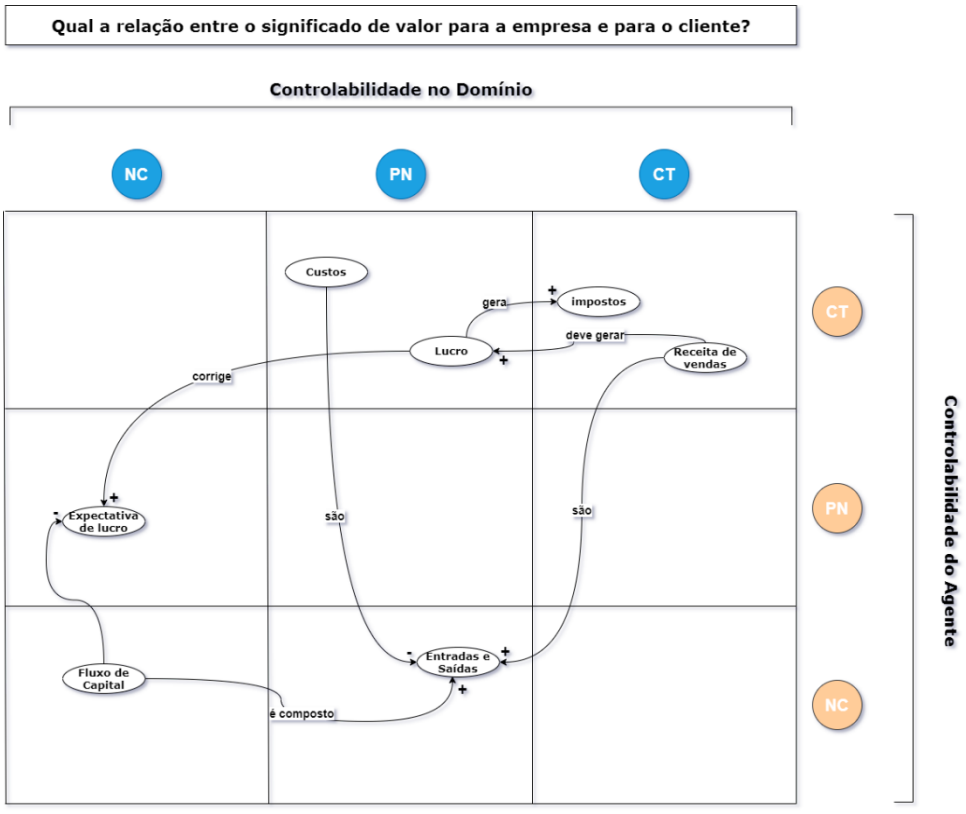
Os MCE dos alunos 1, 2, 3, 4 e do professor são apresentados nas figuras 7.1, 7.2, 7.3, 7.4 e 7.5, respectivamente.

Figura 7.1 – Representação gráfica do MCE do Aluno 1



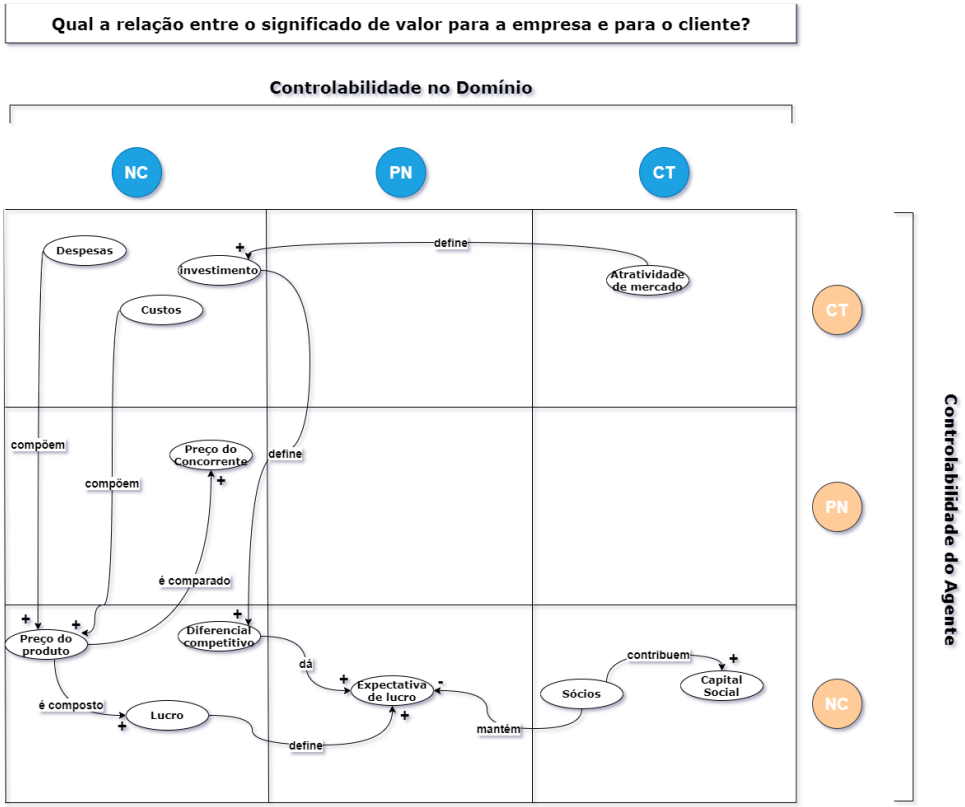
Fonte: Desenvolvido pelo autor

Figura 7.2 – Representação gráfica do MCE do Aluno 2



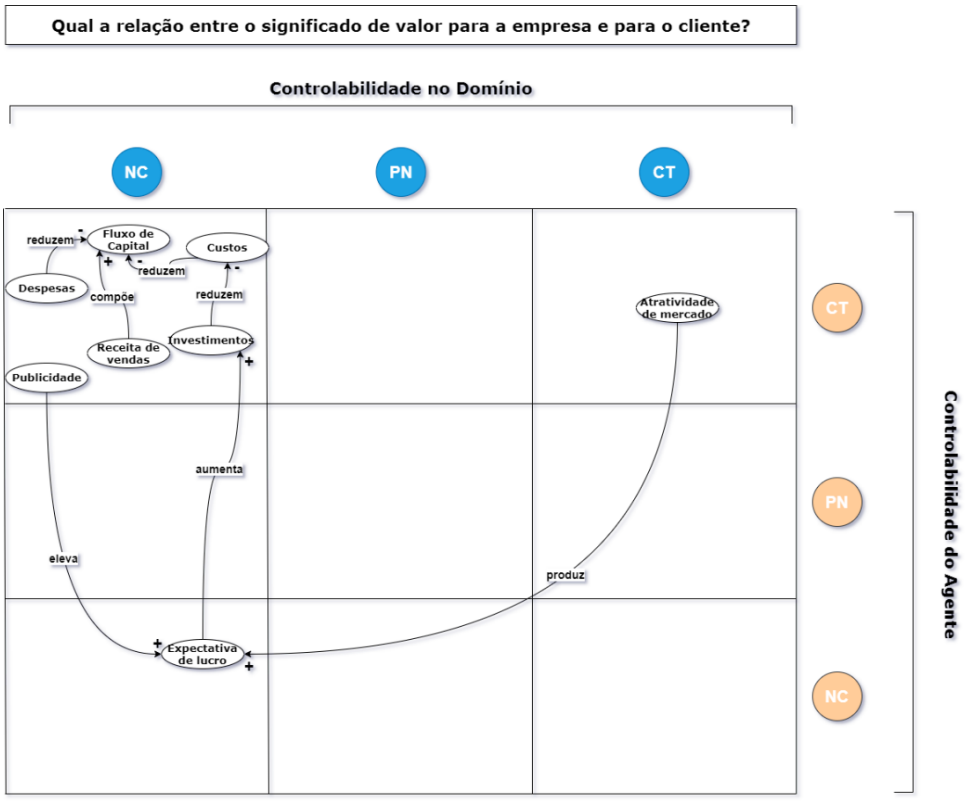
Fonte: Desenvolvido pelo autor

Figura 7.3 – Representação gráfica do MCE do Aluno 3



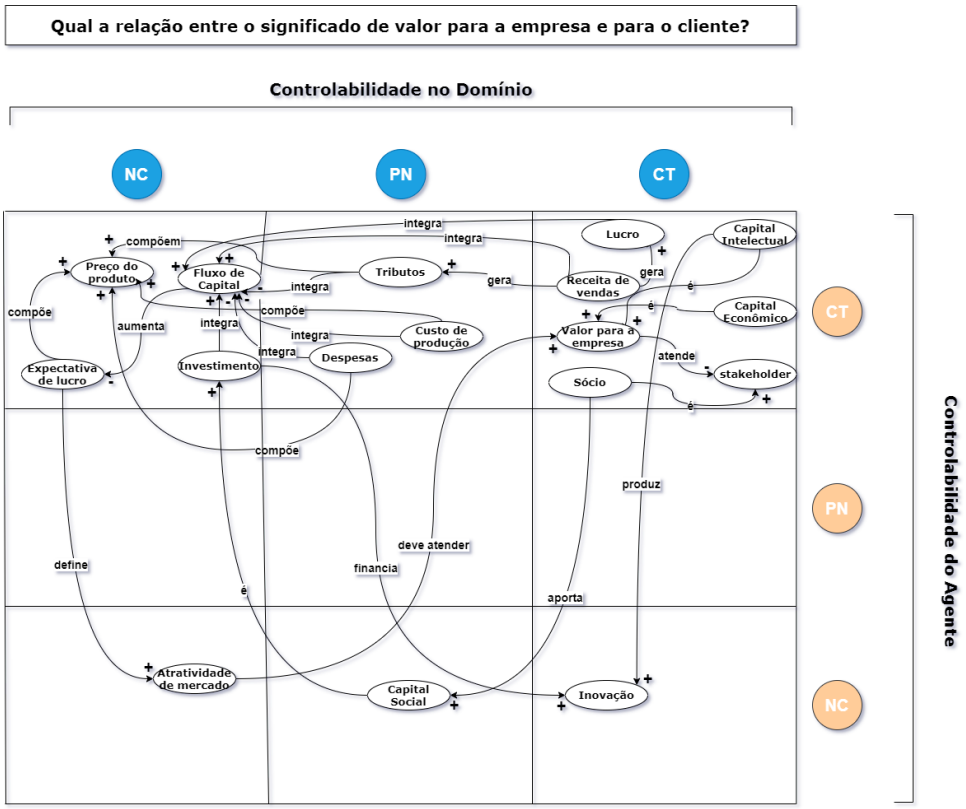
Fonte: Desenvolvido pelo autor

Figura 7.4 – Representação gráfica do MCE do Aluno 4



Fonte: Desenvolvido pelo autor

Figura 7.5 – Representação gráfica do MCE do professor



Fonte: Desenvolvido pelo autor

O mapa do professor (Figura 7.5) foi considerado o MCE de referência, o qual foi comparado com os quatro MCE dos alunos usando o processo de comparação de MCE apresentado no Capítulo 6. Foram consideradas três possibilidades de análise de acordo com as três funções de similaridade (*Fs*) (Capítulo 5, Seção 5.4). Foram utilizados apenas quatro MCE a serem comparados a um MCE de referência, de modo a ser possível comparar os resultados obtidos computacionalmente com uma avaliação visual dos MCE das figuras 7.1, 7.2, 7.3, 7.4 e 7.5, conforme abordado na próxima seção.

7.1 Resultados e discussões

Os resultados das análises são apresentados com os valores de similaridade em ordem decrescente, ou seja, do mais similar para o menos similar em relação ao MCE do professor (MCE de referência).

A Tabela 7.2 mostra os resultados da etapa *(II)* do processo de comparação de MCE (Figura 6.1) que resulta nas melhores correspondências de rótulos de conceitos de dois MCE. Foram utilizados os MCE do professor e do Aluno 3 como exemplo.

Tabela 7.2 – Resultados da similaridade entre conceitos

Conceitos (MCE Prof.)	**Conceitos (MCE Aluno 3)**	**Similaridade**
Preço do Produto	Preço do produto	1.0
Lucro	Lucro	1.0
Investimento	investimento	1.0
Expectativa de lucro	Expectativa de Lucro	1.0
Atratividade de Mercado	Atratividade de Mercado	1.0
Capital Social	Capital Social	1.0
Despesas	Despesas	1.0
Sócio	Sócios	0.9
Custo de Produção	Custo	0.45
Valor para a empresa	Preço do Concorrente	0.3
stakeholder	**sem correspondência**	0
Tributos	**sem correspondência**	0
sem correspondência	Diferencial competitivo	0
Fluxo de Capital	**sem correspondência**	0
Inovação	**sem correspondência**	0
Receita de Vendas	**sem correspondência**	0
Capital Intelectual	**sem correspondência**	0
Capital Econômico	**sem correspondência**	0

Fonte: Desenvolvido pelo autor

A Tabela 7.3 mostra um ranqueamento contendo o grau de similaridade obtido por meio da $Fs(\mathrm{VC1}_{\mathrm{MCE1}}, \mathrm{VC1}_{\mathrm{MCE2}})$ para todos os MCE dos alunos comparados ao MCE do professor. Os valores obtidos podem ser utilizados nas consultas por similaridade para retornar os MCE mais próximos ao MCE do professor, utilizando a similaridade entre conceitos com seus posicionamentos na matriz de atributos (*VC1*).

Tabela 7.3 – Resultados da correspondência entre conceitos

MCE	**Similaridade**
Aluno 3	0.337
Aluno 4	0.287
Aluno 1	0.260
Aluno 2	0.207

Fonte: Desenvolvido pelo autor

Observa-se na Tabela 7.3 que o MCE do Aluno 3 apresenta o maior valor de similaridade e, pela Tabela 7.1, constata-se que contém a maior quantidade de conceitos em relação aos MCE dos outros alunos. Entretanto, a quantidade de conceitos pode interferir positivamente ou negativamente no resultado, pois o cálculo da $Fs(VC1_{\mathrm{MCE1}}, VC1_{\mathrm{MCE2}})$ considera todos os pares de rótulos de conceitos (Tabela 7.2), com correspondência ou não. Ou seja, a utilização de muitos conceitos que estejam fora do contexto da questão focal, contribui negativamente com o resultado da similaridade.

A Tabela 7.4 apresenta o ranqueamento dos alunos contendo o grau de similaridade utilizando a $Fs(VC1_{\mathrm{MCE1}}, VC1_{\mathrm{MCE2}})$ para MC, ou seja, sem o posicionamento dos conceitos.

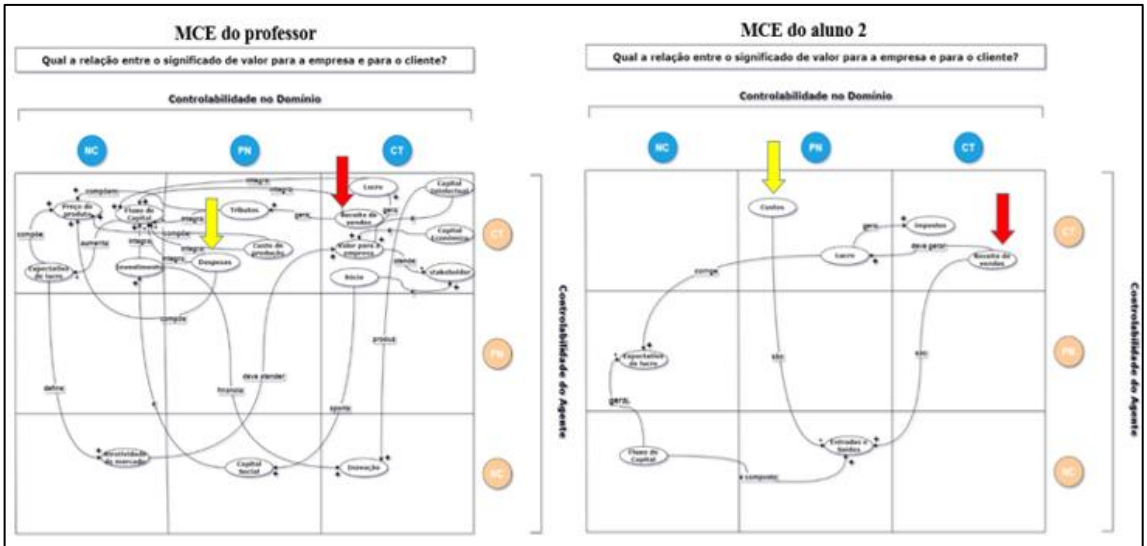
Tabela 7.4 – Resultados da correspondência entre conceitos sem posicionamento

MCE	**Similaridade**
Aluno 3	0.480
Aluno 4	0.352
Aluno 1	0.314
Aluno 2	0.257

Fonte: Desenvolvido pelo autor

É possível notar que removendo o posicionamento dos conceitos a ordem dos alunos é mantida e os valores apresentam um pequeno aumento. Isso pode ser explicado pois a análise fica menos restrita quando desconsiderado o posicionamento dos conceitos. Para ilustrar, a Figura 7.6 apresenta a correspondência de conceitos entre o MCE do professor e do Aluno 2. Nesse caso, houve menor variação de similaridade (0.207 → 0.257), pois, de cinco conceitos correspondidos, dois possuem o mesmo posicionamento, ou seja, estão localizados no mesmo quadrante da matriz de atributos. Por outro lado, quando o MCE do professor é comparado ao MCE do Aluno 3 (Figura 7.7), de dez conceitos correspondidos, apenas um possui o mesmo posicionamento. Desse modo, o MCE do Aluno 3 obteve maior aumento de similaridade (0.337 → 0.480).

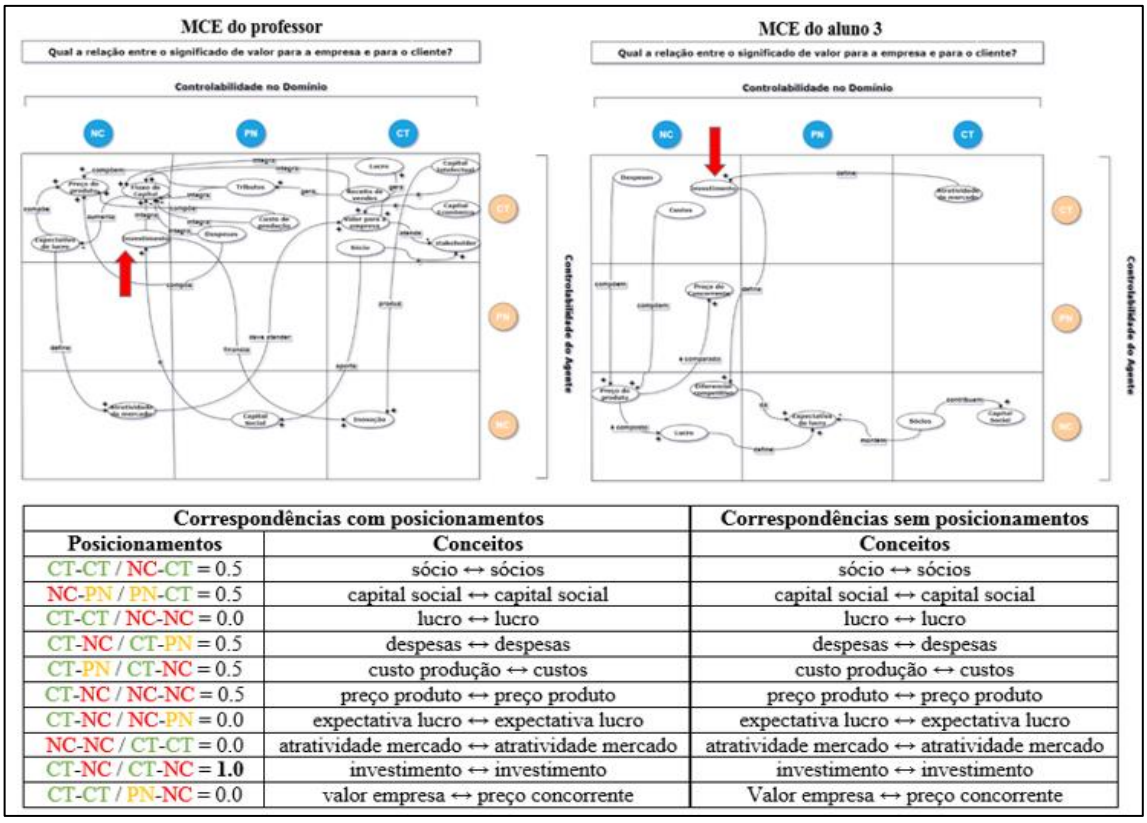
Figura 7.6 – Correspondência de conceitos entre o MCE do professor e do Aluno 2



Correspondências com posicionamentos		Correspondências sem posicionamentos
Posicionamentos	**Conceitos**	**Conceitos**
CT-CT / CT-CT = **1.0**	sócio ↔ sócios	sócio ↔ sócios
CT-CT / CT-PN = 0.5	capital social ↔ capital social	capital social ↔ capital social
CT-PN / CT-PN = **1.0**	lucro ↔ lucro	lucro ↔ lucro
CT-NC / NC-NC = 0.5	despesas ↔ despesas	despesas ↔ despesas
CT-NC / NC-PN = 0.0	custo produção ↔ custos	custo produção ↔ custos

Fonte: Desenvolvido pelo autor

Figura 7.7 – Correspondência de conceitos entre o MCE do professor e do Aluno 3



Correspondências com posicionamentos		Correspondências sem posicionamentos
Posicionamentos	**Conceitos**	**Conceitos**
CT-CT / NC-CT = 0.5	sócio ↔ sócios	sócio ↔ sócios
NC-PN / PN-CT = 0.5	capital social ↔ capital social	capital social ↔ capital social
CT-CT / NC-NC = 0.0	lucro ↔ lucro	lucro ↔ lucro
CT-NC / CT-PN = 0.5	despesas ↔ despesas	despesas ↔ despesas
CT-PN / CT-NC = 0.5	custo produção ↔ custos	custo produção ↔ custos
CT-NC / NC-NC = 0.5	preço produto ↔ preço produto	preço produto ↔ preço produto
CT-NC / NC-PN = 0.0	expectativa lucro ↔ expectativa lucro	expectativa lucro ↔ expectativa lucro
NC-NC / CT-CT = 0.0	atratividade mercado ↔ atratividade mercado	atratividade mercado ↔ atratividade mercado
CT-NC / CT-NC = **1.0**	investimento ↔ investimento	investimento ↔ investimento
CT-CT / PN-NC = 0.0	valor empresa ↔ preço concorrente	Valor empresa ↔ preço concorrente

Fonte: Desenvolvido pelo autor

A Tabela 7.5 apresenta os resultados da comparação de todos os MCE dos alunos em relação ao MCE do professor, considerando os relacionamentos com pesos entre conceitos, por meio da $Fs(VC2_{MCE1}, VC2_{MCE2})$.

Tabela 7.5 – Resultados da correspondência entre relacionamentos

MCE	Similaridade
Aluno 3	0.103
Aluno 2	0.072
Aluno 1	0.036
Aluno 4	0

Fonte: Desenvolvido pelo autor

De acordo com a Tabela 7.5, se o analista restringir sua análise, considerando conceitos similares que se relacionam *(VC2*), o MCE do Aluno 4 passa ser o menos similar. Isso mostra que o Aluno 4 usou conceitos relacionados a questão focal em seu MCE o que o colocou como o segundo mais similar na Tabela 7.4. Entretanto, esses conceitos não estão corretamente relacionados ao contexto da questão focal.

A Tabela 7.6 apresenta resultados de similaridade da $Fs(VC2_{MCE1}, VC2_{MCE2})$ para MC, ou seja, sem considerar os pesos dos relacionamentos e o posicionamento dos conceitos.

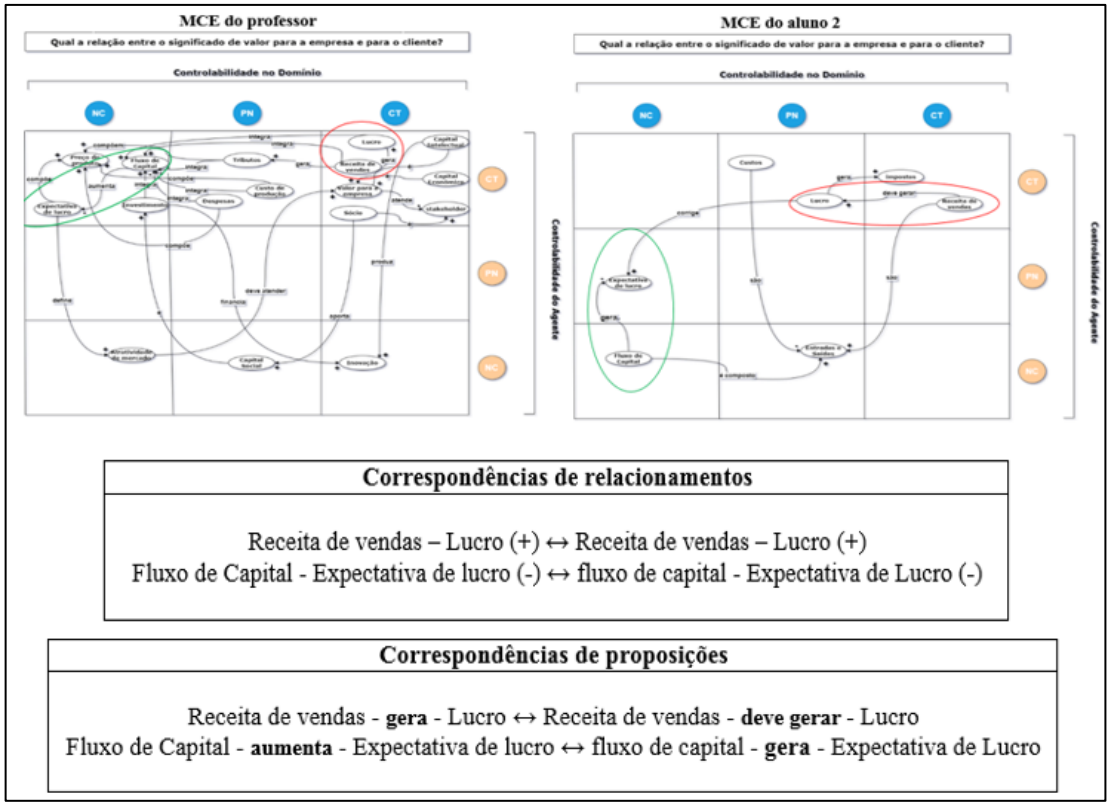
Tabela 7.6 – Resultados da correspondência entre relacionamentos sem posicionamento e pesos

MCE	Similaridade
Aluno 3	0.116
Aluno 2	0.087
Aluno 1	0.043
Aluno 4	0

Fonte: Desenvolvido pelo autor

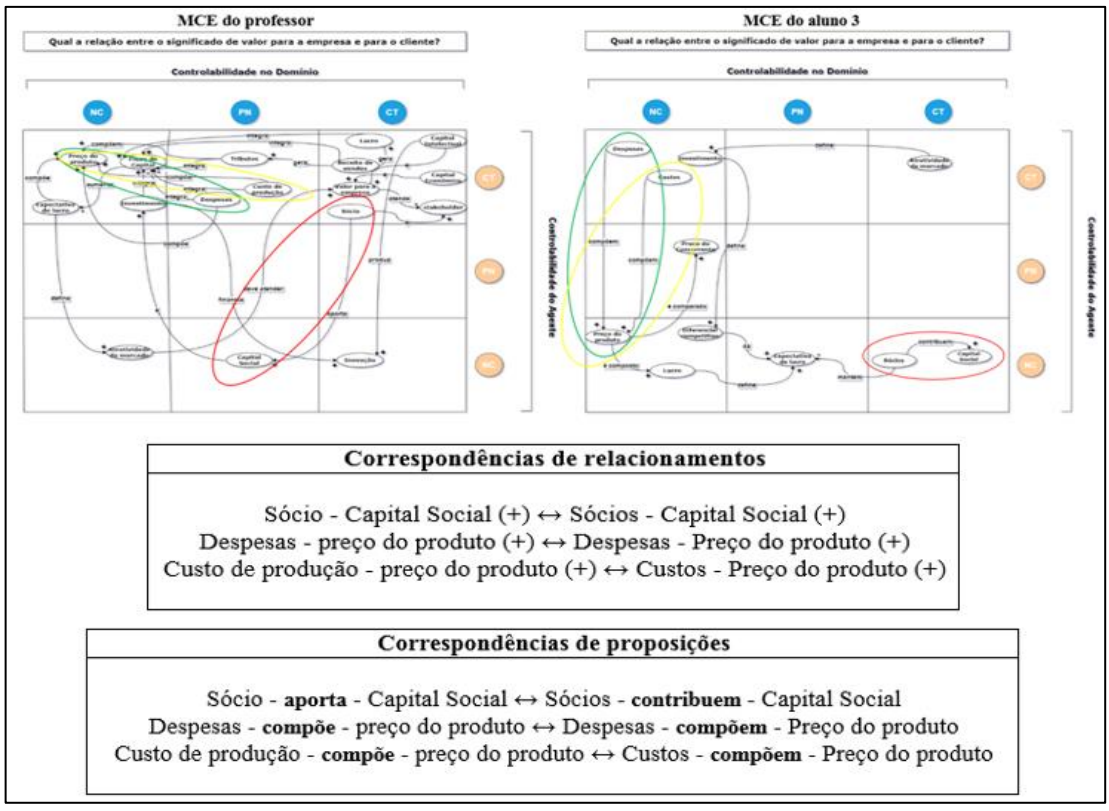
Observando a Tabela 7.6 nota-se uma alteração mínima nos valores de similaridade em relação aos resultados apresentados na Tabela 7.5. As figuras 7.8 e 7.9 apresentam a correspondência de relacionamentos e proposições do MCE do professor com os MCE dos alunos 2 e 3, respectivamente. Observa-se que na parte de correspondência de relacionamentos, os pesos dos relacionamentos são iguais para todas correspondências dos MCE dos alunos 2 e 3 em relação ao MCE do professor. A pequena alteração no valor de similaridade se deu por conta dos posicionamentos dos conceitos.

Figura 7.8 – Correspondência de relacionamentos e proposições entre o MCE do professor e do Aluno 2



Fonte: Desenvolvido pelo autor

Figura 7.9 – Correspondência de relacionamentos e proposições entre o MCE do professor e do Aluno 3



Fonte: Desenvolvido pelo autor

A Tabela 7.7, apresenta os resultados do uso da $Fs(VC3_{MCE1}, VC3_{MCE2})$, onde, considera a similaridade entre as proposições *(VC3)* dos MCE dos alunos em relação ao MCE do professor.

Tabela 7.7 – Resultados da correspondência entre proposições

MCE	Similaridade
Aluno 3	0.097
Aluno 2	0.059
Aluno 1	0.027
Aluno 4	0

Fonte: Desenvolvido pelo autor

Considerando a Tabela 7.7, é possível observar que a ordem dos alunos não se altera em relação a Tabela 7.5. Porém, os resultados de similaridade têm uma pequena queda. Isso significa que, com a adição da similaridade dos termos de ligação aos relacionamentos, restringe-se ainda mais a análise e a acurácia dos resultados da comparação dos MCE tende a aumentar.

A Tabela 7.8 apresenta resultados de similaridade da $Fs(VC3_{MCE1}, VC3_{MCE2})$ para MC, ou seja, sem considerar os pesos dos relacionamentos e o posicionamento dos conceitos.

Tabela 7.8 – Resultados da correspondência entre proposições sem posicionamento e pesos

MCE	Similaridade
Aluno 3	0.103
Aluno 2	0.064
Aluno 1	0.028
Aluno 4	0

Fonte: Desenvolvido pelo autor

Ao analisar os resultados da Tabela 7.8 comparados a Tabela 7.7, tem-se uma conclusão análoga aos resultados obtidos para os relacionamentos, como ilustrado pelas figuras 7.8 e 7.9. A adição da similaridade dos termos de ligação (TL) aos resultados dos relacionamentos contribuiu para uma pequena queda no valor de similaridade das correspondências das proposições. A Tabela 7.9 apresenta os valores de similaridade dos termos de ligação para os alunos 2 e 3 em relação ao MCE do professor. Observa-se que, tanto para o Aluno 2, quanto para o Aluno 3, uma das proposições correspondidas, teve valor de similaridade de seus termos de ligação igual a 0.

Tabela 7.9 – Valores de similaridade dos termos de ligação entre o MCE do professor e dos alunos 2 e 3

MCE Professor x MCE Aluno 2			**MCE Professor x MCE Aluno 3**		
TL Professor	**TL Aluno 2**	**Similaridade**	**TL Professor**	**TL Aluno 3**	**Similaridade**
gera	deve gerar	0.45	aporta	contribuem	0
aumenta	gera	0	compõe	compõem	0.9
			compõe	compõem	0.9

Fonte: Desenvolvido pelo autor

No estudo de caso foram utilizados apenas quatro MCE a serem comparados a um MCE de referência, de modo que é possível comparar os resultados obtidos computacionalmente com uma avaliação visual dos MCE das figuras 7.1, 7.2, 7.3, 7.4 e 7.5.

Os ranqueamentos de valores de similaridade mostrados nas tabelas 7.3, 7.5 e 7.7 podem ser utilizados como base para as consultas por similaridade. Considerando um conjunto maior de MCE, é possível obter os MCE dos alunos cuja similaridade ao MCE do professor seja maior ou igual a um grau de similaridade fornecido, o que exemplifica uma *RQ*. Por exemplo, no caso dos resultados obtidos na Tabela 7.3, o analista poderia escolher verificar visualmente apenas os MCE dos alunos com pelo menos 30% de similaridade em relação ao MCE do professor, o que retornaria apenas os MCE do Aluno 3. Ou então, retornar os *k* MCE dos alunos mais similares ao MCE do professor, em uma *kNNQ*. Por exemplo, considerando os resultados obtidos na Tabela 7.5, é possível obter apenas os 2 MCE ($k = 2$) com maior grau de similaridade em relação ao MCE do professor, resultando nos MCE dos alunos 2 e 3. Desse modo, o uso das consultas por similaridade possibilita ao analista filtrar os resultados, reduzindo o número de mapas para uma análise visual.

7.2 Considerações finais do capítulo

Este capítulo apresentou um estudo de caso para aplicar o processo de comparação de MCE abordado no Capítulo 6, onde os resultados obtidos foram apresentados e discutidos.

8 CONCLUSÃO

8.1 Considerações gerais

Esta dissertação apresentou um processo de comparação por similaridade de conhecimento representado por meio de MCE como uma solução computacional que viabiliza a realização de diferentes tipos de análises. O MCE foi definido como um tipo de dado complexo contendo partes estruturadas e não estruturadas, apresentando-se também o esquema conceitual do banco de dados para armazenamento desse tipo de dado.

Uma vez armazenados em um banco de dados, as características dos MCE podem ser utilizadas em operações de comparação efetuadas para responder consultas por similaridade. Devido a estrutura do MCE, é possível extrair três tipos de vetores de características e, então, definir três funções de similaridade, a serem usadas nas consultas por similaridade de acordo com a análise que o analista julgar necessária.

As comparações por similaridade de MCE realizadas pelas funções de similaridade propostas tomam como base categorias de técnicas de correspondência utilizadas em modelos de representação de conhecimento que mais se assemelham às características do MCE. Para as comparações dos MCE foi proposto o uso de técnicas de correspondência complexas a partir do uso combinado de técnicas de correspondência de diferentes categorias.

O processo de comparação por similaridade não é limitado a um domínio de aplicação específico e considera conceitos e termos de ligação com uma ou mais palavras. Como um MCE é composto por um MC, uma das possibilidades de análise viabiliza a comparação de MC, obtendo-se, assim, um resultado decorrente.

Por meio de um estudo de caso, usando apenas quatro MCE a serem comparados a um MCE de referência, foi possível aplicar o processo proposto e apresentar os resultados obtidos, bem como comparar esses resultados computacionais com uma avaliação visual.

8.2 Trabalhos futuros

Como trabalhos futuros, pretende-se:

- Avaliar o algoritmo em termos de eficiência em outros conjuntos de dados reais e sintéticos, utilizando uma massa de dados significativa e considerando diversos tipos de domínio.

- Integrar as *Fs* propostas a algoritmos de consultas por similaridade, como por exemplo, as consultas por abrangência e as consultas aos *k*-vizinhos mais próximos. Estender a integração a algoritmos de mineração por similaridade, como por exemplo, o *k-medoids*.

- Explorar o uso de MCE em domínios específicos, de modo a verificar se os valores atribuídos aos pesos dos relacionamentos e aos posicionamentos dos conceitos utilizados nas *Fs* possam ser dependentes do domínio. Ou seja, caso isso ocorra, é possível aumentar a acurácia das consultas orientadas a domínios específicos por meio da parametrização dos valores atribuídos aos pesos dos relacionamentos e aos posicionamentos dos conceitos.

REFERÊNCIAS BIBLIOGRÁFICAS

ABBAGNANO, Nicola. **Dicionário de filosofia**. 4. ed. São Paulo: Martins Fontes, 2003.

ABEL, Mara; FIORINI, Sandro R. Uma revisão da Engenharia do Conhecimento: Evolução, Paradigmas e Aplicações. **International Journal of Knowledge Engineering and Management**, v. 2, n. 2, p. 1-35, 2013.

ANGELES, Maria; ESPINO-GAMEZ, Adrian. Comparison of methods Hamming Distance, Jaro, and Monge-Elkan. In: THE SEVENTH INTERNATIONAL CONFERENCE ON ADVANCES IN DATABASES, KNOWLEDGE, AND DATA APPLICATIONS (DBKDA), 2016, Roma, Itália. **Proceedings...** Itália: DBKDA 2016. p. 63-69.

AUSUBEL, David P. **The psychology of meaningful verbal learning**. New York, NY: Grune & Stratton, 1963.

BADDELEY, Alan. Working Memory. **Science**, v. 255, n. 5044, p. 556–559, 1992.

BUNEMAN, Peter. Semistructured Data. In: SIXTEENTH ACM SIGACT-SIGMOD-SIGART SYMPOSIUM ON PRINCIPLES OF DATABASE SYSTEMS, 1997, Tucson, Arizona, USA. **Proceedings...** Arizona: ACM 1997. p. 117-191.

CALDAS, Vanessa M.; FAVERO, Eloi L. Uma Ferramenta de Avaliação Automática para Mapas Conceituais como Auxílio ao Ensino em Ambientes de Educação a Distância. In: TWENTIETH BRAZILIAN SYMPOSIUM ON COMPUTERS IN EDUCATION, 2009, Florianópolis, SC, Brasil. **Anais…** SC: SBIE 2009.

CAO, Mengyun; SUN, Xiaoping; ZHUGE, Hai. The contribution of cause-effect link to representing the core of scientific paper—The role of Semantic Link Network. **PLOS ONE**, v. 13, n. 6, p. 1-14, 2018.

CARD, Stuart K.; MACKINLAY, Jock D.; SHNEIDERMAN, Ben. **Readings in information visualization: using vision to think**. San Francisco, CA: Morgan Kaufmann Pub., 1999.

CHAVÉZ, Edgar; NAVARRO, Gonzalo; BAEZA-YATES, Ricardo; MARROQUÍN, José L. Searching in metric spaces. **ACM Computing Surveys (CSUR)**, v. 33, p. 273-321, 2001.

CHEN, Sei-Wang; LIN, S. C.; CHANG, Kuo E. Attributed Concept Maps: Fuzzy Integration and Fuzzy Matching. **IEEE Transactions on Systems, Man and Cybernetics – Part B (Cybernetics)**, v. 31, n. 5, p. 842-852, 2001.

CLARK, Herbert H. **Using language**. Cambridge, UK: Cambridge University Press, 1996.

COELHO, Alexandre R. **Stemming para a língua portuguesa: estudo, análise e melhoria do algoritmo RSLP**. 2007. 69f. Trabalho de Conclusão de Curso (Ciência da Computação) – Universidade Federal do Rio Grande do Sul, Instituto de Informática, Porto Alegre, RS, 2007.

COHEN, William W.; RAVIKUMAR, Pradeep; FIENBERG, Stephen E. A Comparison of String Distance Metrics for Name-Matching Tasks. In: INTERNATIONAL CONFERENCE ON INFORMATION INTEGRATION ON THE WEB, 2003, Acapulco, México. **Proceedings…** México: IIWEB 2003. p. 73-78.

CRESTANI, Fabio. Application of spreading activation techniques in information retrieval. **Artificial Intelligence Review**, v. 11, p. 453-482, 1997.

CRUSE, David A. **Lexical semantics**. Cambridge, UK: Cambridge University Press, 1986

DA SILVA, Bento C. D. Wordnet.br: An Exercise of Human Language Technology Research. In: THIRD INTERNATIONAL WORDNET CONFERENCE, 2006, Seogwipo, Korea. **Proceedings…** Korea: GWC 2006. p. 301-303.

DE SOUZA, Flávio S. L.; BOERES, Maria C. S.; BOERES, D.; DE MENEZES; C. S.; CARLESSO, G. An Approach to Comparison of Concept Maps Represented by Graphs. In: THIRD INTERNATIONAL CONFERENCE ON CONCEPT MAPPING, 2008, Tallinn, Estonia and Helsinki, Finland. **Proceedings…** Estonia and Finland: CMC 2008. p. 205-212.

DUARTE, Gabriel A. **Visualização de mapas conceituais estendidos utilizando grafos orientados a força e restrições de posicionamento de nós**. 2018. 97f. Dissertação (Tecnologia e Inovação) – Universidade Estadual de Campinas, Faculdade de Tecnologia, Limeira, SP, 2018.

ELMASRI, Ramez; NAVATHE, Shamkant B. **Fundamentals of database systems**. Boston, MA: Addison-Wesley, 2011.

FREEMAN, Andrew; CONDON, Sherri; ACKERMAN, Christopher. Cross Linguistic Name Matching in English and Arabic: A "One to Many Mapping" Extension of the Levenshtein Edit Distance Algorithm. In: ANNUAL CONFERENCE OF THE NORTH AMERICAN CHAPTER OF THE ASSOCIATION FOR COMPUTATIONAL LINGUISTICS, 2006, Boston, USA. **Proceedings...** USA: HTL-NAACL 2006. p. 471-478.

GOLDSMITH, Timothy E.; JOHNSON, Peder J.; ACTON, William H. Assessing Structural Knowledge. **Journal of Educational Psychology**, v. 83, p. 88-96, 1991.

GOMES, Franciene D. **Ferramentas de Gestão do Conhecimento e Estilos de Aprendizagem para Apoio às Estratégias Pedagógicas no Ensino Superior**. 2018. 146f. Dissertação (Doutorado em Tecnologia) - Universidade Estadual de Campinas, Faculdade de Tecnologia, Limeira, SP, Brasil, 2018.

GRIMM, Stephan; HITZLER, Pascal; ABECKER, Andreas. Knowledge Representation and Ontologies. In: GRIMM, **Semantic Web Services**, Heidelberg, Germany: Springer, 2007. cap. 3, p. 51-105.

GRUBER, Thomas R. Toward Principles for the Design of Ontologies Used for Knowledge Sharing. **International Journal of Human-Computer Studies**, v. 43, n. 5-6, p. 907-928, 1995.

HE, Wei; YANG, Xiaoping; HUANG, Dupei. A hybrid approach for measuring semantic similarity between ontologies based on wordnet. In: KNOWLEDGE SCIENCE,

ENGINEERING AND MANAGEMENT, 2011, Irvine, CA. **Proceedings...** CA: Springer, Berlin, Heidelberg 2011. p. 68-78.

HJORLAND, Birger. Concept theory. **Journal of the American Society for Information Science and Technology**, v. 60, n. 8, p. 1519-1536, 2009.

IFENTHALER, Dirk. Relational, structural, and semantic analysis of graphical representations and concept maps. **Educational Technology Research and Development**, v. 58, p. 81–97, 2008.

JARAMILLO, Juan F. G. **Identificação da Percepção de Valor Público por Meio de Ferramentas de Gestão do Conhecimento**. 2018. 85f. Dissertação (Mestrado em Tecnologia) - Universidade Estadual de Campinas, Faculdade de Tecnologia, Limeira, SP, Brasil, 2018.

JARO, Matthew. A. Probabilistic linkage of large public health data files. **Statistics in Medicine**, v. 15, n. 5-7, p. 491-498, 1995.

JOSLYN, Cliff A.; PAULSON, P.; WHITE, Amanda. Measuring the structural preservation of semantic hierarchy alignment, In: FOURTH INTERNATIONAL CONFERENCE ON ONTOLOGY MATCHING, 2009, Chantilly, USA. **Proceedings...** USA: OM'09 2009. p. 61-72.

LEHMANN, Fritz. Semantic networks. **Computers & Mathematics with Applications**, v. 23, n. 2-5, p. 1-50, 1992.

LIMONGELLI, Carla; LOMBARDI, Matteo; MARANI, Alessandro; SCIARRONE, Filippo; TEMPERINI, Marco. Concept Maps Similarity Measures for Educational Applications. In: INTERNATIONAL CONFERENCE ON INTELLIGENT TUTORING SYSTEMS, 2016, Zagreb, Croatia. **Proceedings…** Zagreb: ITS 2016. p. 361-367.

MARINOV, Milko; ZHELIAZKOVA, Irina. An interactive tool based on priority semantic networks. **Knowledge-Based Systems**, v. 18, n. 2-3, p. 71-77, 2005.

MARKOWITZ, Judith A.; NUTTER, Terry J.; EVENS, Martha W. Beyond IS-A and part-whole: More semantic network links. **Computers & Mathematics with Applications**, v. 23, n. 6-9, p. 377-390, 1992.

MARSHALL, Byron; CHEN, Hsinchun; MADHUSUDAN, Therani. Matching knowledge elements in concept maps using a similarity flooding algorithm. **Decision Support Systems**, v. 42, n. 3, p. 1290-1306, 2006.

MARTINS, Eliane; MACHADO, Gustavo. Análise do Uso de Diversas Funções de Similaridade no Agrupamento de Casos de Teste Baseados em Modelos de Estado. In: SIMPÓSIO BRASILEIRO DE REDES DE COMPUTADORES E SISTEMAS DISTRIBUÍDOS (SBRC), 2018, Campos do Jordão – SP, Brasil. **Anais...** Campos do Jordão: SBRC, WTF 2018. p. 1–13.

MASCARDI, Viviana; LOCORRO; Angela, ROSSO; Paolo. Automatic Ontology Matching via Upper Ontologies: A Systematic Evaluation. **IEEE Transactions on Data Engineering**, v. 22, n. 1, p. 609-623, 2009.

MELO, Maria A. F.; BRÄSCHER, Marisa. Termo, conceito e relações conceituais: um estudo das propostas de Dahlberg e Hjorland. **Ciência da Informação**, v. 43, n. 1, p. 67–80, 2015.

MONTES-Y-GÓMES, Manuel; LÓPEZ-LÓPEZ, Aurelio; GELBUKH, Alexander. Information retrieval with conceptual graph matching. In: INTERNATIONAL CONFERENCE ON DATABASE AND EXPERT SYSTEMS APPLICATIONS, 2000, Heidelberg, Berlin. **Proceedings...** Berlin: DEXA 2000. p. 312-321.

MOREIRA, Antônio L. M.; HAYASHI, Thomas; COELHO, Guilherme P.; SILVA, Ana E. A. A Clustering Method for Weak Signals to Support Anticipative Intelligence. **International Journal of Artificial Intelligence and Expert Systems (IJAE)**, v. 6, p. 1-14, 2015.

MOREIRA, Marco A.; BUCHWEITZ, Bernardo. **Mapas Conceituais: Instrumentos didáticos de Avaliação de análise de currículos**. 1. ed. São Paulo: Moraes, 1987.

MOREIRA, Marco. A. **Mapas Conceituais e Aprendizagem Significativa**. Porto Alegre: Instituto de Física, UFRGS, 1998, 10p (adaptado e atualizado, em 1997, de MOREIRA, Marco. A. Mapas Conceituais e Aprendizagem Significativa. **O Ensino**, Pontevadra/Espanha e Braga/Portugal, n. 23 a 28, p. 87–95, 1988)

MULLER, Henning; MICHOUX, Nicolas; BANDON, David; GEISSBUHLER, Antoine. A review of content-based image retrieval systems in medical applications-clinical benefits and future directions. **International Journal of Medical Informatics (IJMI)**, v. 73, n. 1, p. 1-23, 2004.

MURPHY, Gregory. **The big book of concepts**. Cambridge, Massachusetts: The MIT Press, 2002.

NEGHEVITSKY, M. **Artifical Intelligence: A Guide to Intelligent Sistems**. 2. ed., Harlow, England: Addison-Wesley, 2005.

NONAKA, Ikujiro; TAKEUCHI, Hirotaka; UMEMOTO, Katsuhiro. A theory of organizational knowledge creation. **International Journal of Technology Management (IJTM) Special Publication on Unlearning and Learning**, v. 11, n. 7-8, p. 833-845, 1996.

NONAKA, Ikujiro. A Dynamic Theory of Organizational Knowledge Creation. **Organizational Science**, v. 5, n. 1, p. 14-37, 1994.

NOVAK, Joseph D.; CAÑAS, Alberto J. The Theory Underlying Concept Maps and How to Construct and Use Them. **IHMC CmapTools**, p. 1-36, 2008.

NOVAK, Joseph D. **Learning, creating, and using knowledge: Concept maps as facilitative tools in schools and corporations**. 2. ed. Abingdon: Routledge, 2010.

NOVAK, Joseph. D.; MUSONDA, Dismas. A twelve-year longitudinal study of science concept learning. **American Educational Research Journal**, v. 28, n. 1, p. 117-153, 1991.

NOY, Natalya F.; HAFNER, Carole D. The State of the Art in Ontology Design. **AI Magazine**, v. 18, n. 3, p. 53-74, 1997.

OLIVEIRA, Vitor M.; Roman, Norton T. JWN-Br – an Java API for WordNet.Br. In: NINETH BRAZILIAN SYMPOSIUM IN INFORMATION AND HUMAN LANGUAGE TECHNOLOGY, 2013, Fortaleza, Ceará, Brasil. **Proceedings...** Ceará: STIL 2013. p. 153-157.

ORENGO, Viviane M.; HUCYK, Christian. A stemming algorithm for the portuguese language. In: EIGHTH SYMPOSIUM ON STRING PROCESSING AND INFORMATION RETRIEVAL, 2001, Laguna de San Rafael, Chile. **Proceedings...** Chile: IEEE 2001. p. 186-193.

OTERO-CERDEIRA, Lorena; RODRÍGUEZ-MARTÍNEZ, Francisco J.; GÓMEZ-RODRÍGUEZ, Alma. Ontology matching: A literature review. **Expert Systems with Applications**, v.42, n. 2, p. 949-971, 2015

PÉREZ, Cláudia C. C.; VIEIRA, Renata. Mapas Conceituais: geração e avaliação. In: III WORKSHOP EM TECNOLOGIA DA INFORMAÇÃO E DA LINGUAGEM HUMANA (TIL), 2005, São Leopoldo-RS, Brasil. **Anais...** São Leopoldo: TIL, UNISINOS 2005. p. 2158-2167.

PIRES, Carlos E.; SOUZA, Damires; PACHÊCO, SALGADO; Ana C. SemMatcher: A Tool for Matching Ontology-based Schemas. In: TWENTY-FOURTH BRAZILIAN SYMPOSIUM ON DATA BASES (SBBD'09), 2009, Fortaleza, Brasil. **Anais…** Fortaleza: SBBD'09 2009.

PRIEST, Graham. **Logic: A Very Short Introduction**. New York, USA: Oxford University Press, 2001.

QUILLIAN, Ross M. Semantic Memory. In: QUILLIAN, **Semantic Information Processing**. M. Minsky, Cambridge: MIT Press, 2003. p 227-270.

RUIZ-PRIMO, Maria A.; SHAVELSON, Richard J. Problems and issues in the use of Concept Maps in science assessment. **Journal of Research in Science Teaching**, v. 33, n. 6, p. 569–600, 1996.

SAFAYENI, Frank; DERBENTSEVA, Natalia; CAÑAS, Alberto J. A theoretical note on concepts and the need for cyclic concept maps. **Journal of Research in Science Teaching**, v. 42, n. 7, p. 741–766, 2005.

SAFAYENI, Frank; DERBENTSEVA, Natalia; CAÑAS, Alberto J. Concept Maps: Experiments on Dynamic Thinking. **Journal of Research in Science Teaching**, v. 44, n. 3, p. 448-465, 2006.

SAMET, Hanan. **Foundations of Multidimensional and Metric Data Structures**. 1. ed. San Francisco, CA: Morgan Kaufmann Publishers, 2006.

SAPHIRO, Stuart C. SNePS: A Logic for Natural Language Understanding and Commonsense Reasoning. In: SAPHIRO, **Natural Language Processing and Knowledge representation: Language for Knowledge and Knowledge for Language**. Cambridge: MIT Press, 2000. p. 175-195.

SENGE, Peter M. **The fifth discipline fieldbook: Strategies and tools for building a learning organization**. 1. ed. New York, NY: Crown Business, 2014.

SHAH, Gauri; SYEDA-MAHMOOD, Tanveer. Searching databases for sematically-related schemas. In: TWENTY-SEVENTH ANNUAL INTERNATIONAL ACM SIGIR CONFERENCE ON RESEARCH AND DEVELOPMENT IN INFORMATION RETRIEVAL, 2004, Sheffield, UK. **Proceedings...** UK: ACM 2004. p. 504-505.

SILBERCHATZ, Abraham; KORTH, Henry; SUDARSHAN, S. Complex Data Types. In: SILBERCHATZ, **Database System Concepts**. 7. ed. New York: McGraw-Hill, 2019. cap. 8, p. 2-37.

SILVA, Alex A; PADILHA, Natália; SIQUEIRA, Sean W. M.; REVOREDO, K; BAIÃO, Fernanda A. Avaliação da aprendizagem apoiada na utilização de mapas conceituais e alinhamento de ontologias. In: WORKSHOP BRASILEIRO DE WEB SEMÂNTICA E EDUCAÇÃO, 2015, Aracaju, Sergipe, Brasil. **Anais…** Sergipe: SWEd 11 2011.

SMITH, Temple F.; WATERMAN, Michael S. Identification of Common Molecular Subsequences. **Journal of Molecular Biology**, v. 14, n. 1, p. 195-197, 1981.

SOWA, John F. **Knowledge Representation: Logical, Philosophical, and Computational Foundations. Pacific Grove**. CA, USA: Brooks Cole Pub. Co., 2000.

STERNBERG, Robert J.; STERMBERG, Karim. The Landscape of Memory: Mental Images, Maps, and Propositions. In: STERNBERG, **Cognitve Psychology**. 6. ed. Harlow, England: Addison-Wesley, 2005. cap. 7, p. 270-318.

TAN, Pang-Ning; STEINBACH, Michael; KUMAR, Vipin. **Introduction to Data Mining**. 1. ed. London, UK: Pearson Education, 2006.

TVERSKY, Amos. Features of similarity. **Psychological Review**, v. 84, n. 4, p. 327–352, 1977.
WIELINGA, Bob J. Reflections on 25+ years of knowledge acquisition. **International Journal of Human-Computer Studies**, v. 71, n. 2, p. 211-215, 2013.

VEIKIRI, Ioanna. What Is the Value of Graphical Displays in Learning?. **Educational Psychology Review**, v. 14, n. 3, p. 261-312, 2002.

WINKLER, William E. Overview of Record Linkage and Current Research Directions. Statistical Research Division. **Relatório**. Washington, 2006.

ZAMBON, Antonio C.; BAIOCO, Gisele B.; CHISTE, Cristiano; VASQUES, Dildre G. Uma aplicação prática de gestão do conhecimento e simulação na resolução de problemas complexos empresariais. **Revista Produção Online**, v. 16, n. 2, p. 408–440, 2016a.

ZHANG, Junsheng; SUN, Y. An Analogy reasoning model for Semantic Link Network. **International Journal of Digital Content Technology and its Applications**, v. 4, n. 12, p 128-139, 2010.

ZHUGE, Hai. Communities and Emerging Semantics in Semantic Link Network: Discovery and Learning. **IEEE Transactions on Knowledge and Data Engineering**, v. 21, n. 1, p. 785-799, 2009.

ZHUGE, Hai. The Semantic Link Network. In: ZHUGE, **The Knowledge Grid: Toward CyberPhysical Society**, London: World Scientific Pub. Co., 2012. cap. 2, p. 71-233.