

FRAMEWORK PARA FORMAÇÃO E ANÁLISE DE AGRUPAMENTOS DE ESTUDANTES A PARTIR DE SIMILARIDADE SEMÂNTICA COGNITIVA

Rodrigo Ruy Boguski¹

Davidson Cury²

Camila Zacché de Aguiar³

RESUMO: Muitas abordagens de ensino-aprendizagem dependem diretamente da formação de agrupamentos e essa tarefa pode se tornar demasiadamente complexa a depender de uma série de fatores. Esta pesquisa propõe um *framework* como ferramenta para professores, que possibilita realizar agrupamentos de estudantes baseados na similaridade semântica de mapas conceituais por eles produzidos. A abordagem consiste na leitura automatizada dos mapas e sua representação como vetores multidimensionais de características que consideram o contexto. Esses vetores podem ser comparados por redes neurais artificiais a fim de identificar semântica sem a necessidade de estruturas de suporte como modelos pré-definidos, ontologias ou web semântica que, por vezes, sobrecarregam a representação da informação e demandam conhecimento específico sobre um domínio. Para que o *framework* pudesse ser instanciado, produzimos um ambiente de software que o suportasse em atividades pedagógicas. Os resultados de experimentos em sala de aula apontam uma correspondência acima de 80% entre os grupos semânticos propostos por professores e o *framework*, e o confirmam como relevante para auxiliar professores a diminuir a carga cognitiva do processo de aprendizagem a partir do entendimento dos grupos e da proximidade semântica de seus integrantes.

PALAVRAS-CHAVE: Agrupamentos Semânticos; Similaridade Semântica Cognitiva; Mapas Conceituais.

¹ Mestre em Informática pela Universidade Federal do Espírito Santo (UFES). Doutorando em Ciência da Computação pela Universidade Federal do Espírito Santo (UFES). Departamento de Informática, Universidade Federal do Espírito Santo, Av. Fernando Ferrari, 514 - Goiabeiras, Vitória - ES | CEP 29075-910. rodrigoboguski@gmail.com. +55 27 4009-2654; +55 27 4009-2850 (fax).

² Pós-doutor em Educação pela Faculdade de Educação da Universidade Federal do Rio Grande do Sul (FACED UFRGS). Professor na Universidade Federal do Espírito Santo (UFES). Departamento de Informática, Universidade Federal do Espírito Santo, Av. Fernando Ferrari, 514 - Goiabeiras, Vitória - ES | CEP 29075-910 dedecury@gmail.com. +55 27 4009-2654; +55 27 4009-2850 (fax).

³ Doutora em Ciência da Computação pela Universidade Federal do Espírito Santo (UFES). Professora na Universidade Federal do Espírito Santo (UFES). Departamento de Informática, Universidade Federal do Espírito Santo, Av. Fernando Ferrari, 514 - Goiabeiras, Vitória - ES | CEP 29075-910 camila.zacche.aguiar@gmail.com. +55 27 4009-2654; +55 27 4009-2850 (fax).

ABSTRACT: Many teaching and learning approaches heavily depend on forming groupings, a task that can become quite complex due to various factors. This research proposes a framework designed as a tool for teachers, allowing them to group students based on the semantic similarity of concept maps they have created. The approach involves the automated analysis of these maps and their representation as multidimensional feature vectors that account for context. These vectors can then be compared using artificial neural networks to identify semantic similarities, eliminating the need for supporting structures such as predefined models, ontologies, or semantic web technologies, which can sometimes complicate information representation and require specific domain knowledge. To implement the framework, we developed a software environment to support its use in educational activities. Classroom experiments show an over 80% match between the semantic groupings proposed by teachers and those generated by the framework, demonstrating its effectiveness in helping teachers reduce cognitive load by providing insights into group compositions and the semantic proximity of group members.

KEY-WORDS: Semantic Clusters; Cognitive Semantic Similarity; Concept Maps.

1. INTRODUÇÃO

A formação de agrupamentos é uma tarefa central em muitas abordagens de ensino-aprendizagem e essa tarefa pode se tornar demasiadamente complexa e demorada a depender de uma série de fatores como número de participantes, propósito e critérios de agrupamento. Existem diferentes formas para realizar agrupamentos dependendo da finalidade pretendida. Uma estratégia interessante é a formação de grupos que possuam similaridade semântica cognitiva pois permite um tratamento adequado e eficiente para o ensino e aprendizado, ajustando métodos e materiais conforme os perfis e necessidades dos alunos. Essa estratégia facilita a personalização do ensino ao permitir que educadores adaptem abordagens pedagógicas, promovendo um aprendizado mais adequado e eficaz, oferecendo intervenções direcionadas para grupos com dificuldades, enquanto desafiam alunos mais avançados. Nesse contexto, apresentamos um *framework* para formação e análise de grupos de estudantes que possuam similaridade semântica, sendo realizados a partir de mapas conceituais por eles produzidos. Os mapas, enquanto valiosa ferramenta para representação do arcabouço conceitual do aprendiz, subsidiam todas as fases, garantindo facilidade para estruturar conhecimento em forma de proposições e permitir comparações. Os métodos propostos no *framework*, tanto para formação de agrupamentos semânticos quanto para sua análise, podem ajudar alunos e professores a diminuir a carga cognitiva do processo de

aprendizagem a partir da melhor identificação e entendimento de grupos semânticos formados, permitindo detectar dispersões na comunicação, perceber as semelhanças e diferenças cognitivas entre diferentes grupos de alunos, além de auxiliar o professor no planejamento de interações significativas entre os alunos. Este trabalho está organizado da seguinte forma: na Seção 2 apresentamos o referencial teórico; na Seção 3, as técnicas e modelos utilizados para os agrupamentos, na Seção 4, o *framework*, na Seção 5, os experimentos e resultados obtidos e na Seção 6, as conclusões.

2. REFERENCIAL TEÓRICO

É possível agrupar estudantes sob diferentes perspectivas. Algumas técnicas são baseadas na formação de grupos em torno de centroides. Em outras, os grupos são criados seguindo uma hierarquia existente nos participantes (HAN, KAMBER, & PEI, 2011). Em ambos os casos é preciso que sejam definidas características dos indivíduos.

Agrupamentos baseados em características pessoais pré-definidas (OTHMANE & ALI, 2016), (MANSKE & HOPPE, 2017) ou perfis (CIUBUC, DASCALU, TRAUSANMATU, & MARHAN, 2013), (VINELLA, LYKOURTZOU, & MASTHOFF, 2021) classificam entidades a partir de atributos ou qualidades que elas possuem, porém, apesar de relevantes para formar agrupamentos semânticos, apresentam uma dimensão semântica limitada, uma vez que necessitam mapear, antecipadamente, os atributos dos participantes.

Outros trabalhos com essa finalidade abordaram agrupamentos sob a perspectiva de web semântica (OUNNAS, DAVIS, & MILLARD, 2008), abordagens visuais (BODENREIDER & MCCRAY, 2003), redes semânticas (PETERS & SHROBE, 2003), modelos ontológicos, (SIMÓN, LUIGI, & ROSETE, 2007), níveis de engajamento de estudantes (MOUBAYED, INJADAT, SHAMI, & LUTFIYYA, 2020) e aspectos sintáticos de mapas conceituais (LOVATI, CURY, AGUIAR, & AZEREDO, 2017). Essas abordagens, embora tenham sua contribuição, não consideram o conhecimento cognitivo semântico que os integrantes do grupo possuem.

Os agrupamentos por meio de ontologias são realizados a partir da comparação de elementos representados por um modelo de domínio do qual participam (THANGARAJ & MEENATCHI, 2015), (MUHAMMAD & NORDIN, 2013). Essa abordagem é satisfatória para mapear domínios específicos do conhecimento, mas como técnica para agrupamentos, necessita que o elemento a ser comparado satisfaça atributos e regras do modelo definido.

Os agrupamentos que utilizam web semântica (SANTOS, 2009), (ZHUGE, 2009) possuem como restrição natural a imersão ou pertencimento de elementos a ambientes web, além disso, é necessário que haja um consenso sobre um modelo predeterminado relativo ao ecossistema em questão.

Agrupamentos baseado em redes neurais (MIKOLOV, CHEN, CORRADO, & DEAN, 2013), (LE & MIKOLOV, 2014), (JACOB DEVLIN, 2018), (DRAGOMIR, D., & M., 2020), (SHAVLIK & MCNELIS, 2012), usam algoritmos para poder reconhecer padrões escondidos e correlações em dados brutos, agrupar, classificar, aprender e melhorar continuamente com o tempo. O uso de redes neurais de baixa dimensionalidade tem produzido bons resultados para obtenção de semântica sobre conteúdo, mas é preciso realizar um treinamento e sintonia de parâmetros da rede. Elas têm se mostrado como a técnica mais proeminente, adaptável a cenários, eficaz e eficiente para processamento de informação e obtenção de conhecimento. Por este motivo, fundamentamos nosso *framework* nesta área de pesquisa.

3. AGRUPAMENTOS SEMÂNTICOS SOBRE MAPAS CONCEITUAIS

O primeiro passo para realizar agrupamentos semânticos é a obtenção da informação a ser comparada. A fim de realizar agrupamentos semânticos que considerassem, exclusivamente, o conhecimento dos alunos, utilizamos os mapas conceituais (Figura 1) (NOVAK & GOWIN, 1984) como meio para representação da informação.

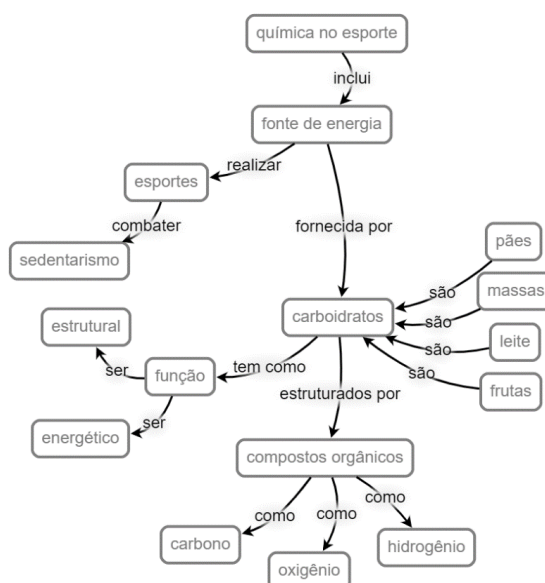


Figura 1. Exemplo de mapa conceitual

Os mapas conceituais são amplamente utilizados em diversas atividades pedagógicas, sendo uma ferramenta eficaz para promover a organização, visualização e compreensão de conceitos, ajudando tanto professores quanto alunos a estruturarem o conhecimento de forma clara, mostrando a relação entre diferentes ideias e explicitando o arcabouço conceitual do aluno. Eles estruturam graficamente conceitos e relações na forma de proposições, permitindo explicitar de forma objetiva quais conceitos estão relacionados e aqueles mais importantes na percepção do aprendiz.

Embora a representação diagramática dos mapas conceituais permita uma leitura e compreensão mais rápidas quando comparadas a um texto que possa tê-lo originado, muitas vezes ocorrem erros em sua construção que podem dificultar a leitura e compreensão das informações descritas, como caminhos em *loop*, ambiguidades conceituais, estruturação não hierárquica ou ainda, uso confuso de frases inteiras em vez de proposições. Apesar de isso ser suficiente para representação da informação, gera muitas dificuldades de compreensão, quer seja para uma pessoa ou mesmo para processamento computacional. Sua representação conceitual, por mais simples que seja, pode ser um pouco confusa até que um certo nível de qualidade seja alcançado. Dessa forma, buscamos uma estratégia que pudesse superar essas dificuldades, sem, no entanto, exigir um treinamento extensivo dos alunos para construir os mapas, permitindo que eles construam mapas melhores a seu tempo.

Existem alguns trabalhos relacionados à leitura e comparação de mapas conceituais (LAMAS, BOERES, CURY, & MENEZES, 2005), (LOVATI, CURY, AGUIAR, & AZEREDO, 2017), (LIMONGELLI, SCIARRONE, LOMBARDI, MARANI, & TEMPERINI, 2017), contudo, a abordagem utilizada reside na comparação das proposições de forma individual ou pela formação de n-gramas. O método aqui proposto, realiza a leitura e comparação de forma mais ampla, considerando o conjunto de conceitos e relações, sua frequência de ocorrência e o relacionamento com os outros conceitos. Isso gera uma correspondência com espectro mais abrangente que a comparação de pares de proposições, resultando na explicitação natural de um assunto.

3.1. LEITURA DOS MAPAS CONCEITUAIS

Desenvolvemos um método de leitura dos mapas que percorre sua estrutura hierárquica e extrai, em forma de texto, a informação nele representada graficamente. A leitura é iniciada com a identificação de nós iniciais e finais. Os nós iniciais são aqueles que não possuem arestas chegando neles, apenas saindo, sendo, geralmente, conceitos de alta hierarquia e funcionando como subsunçores, ou âncoras, para outros conceitos (MOREIRA & RODRÍGUEZ, 1997). Nós finais são aqueles que possuem apenas arestas de entrada. Com a lista de nós iniciais, elege-se, dentre eles, um nó raiz inicial. Um nó raiz é aquele que possui grande relevância no mapa e mais alta hierarquia. Uma sentença é gerada quando se percorre um caminho contínuo entre um nó inicial e nó final sem passar por *loops* no grafo. O texto do mapa é obtido a partir do produto cartesiano $P(\text{nó inicial}, \text{nó final})$.

Embora a leitura seja feita no sentido do nó inicial para o final, ela é dividida em estágios para evitar a repetição de sentenças decorrentes da própria estrutura dos mapas conceituais. Estruturalmente, um nó pode derivar diferentes ramos, a partir de disjunções e junções. Nesses casos, verificamos se os relacionamentos que ligam o nó inicial aos nós finais são iguais ou não. Caso sejam, não devem ser repetidos na construção de cada sentença. Para manter um contexto semântico e respeitar a memória de curto prazo, as sentenças são formadas a partir da ordem de leitura, assim, conceitos são exibidos na sequência em que aparecem na hierarquia. O texto resultante é livre de sentenças redundantes decorrentes de construções em *loop* ou repetição de relacionamentos “um para muitos”, portanto, é organizado de forma semelhante a um texto que possa ter originado o mapa, contudo, em sua versão essencial. O resultado final da leitura do mapa da figura 1 é: “Química no esporte inclui fonte de energia para realizar esportes e combater sedentarismo. Química no esporte inclui fonte de energia fornecida por carboidratos. Pães, massas, leite e frutas são carboidratos. Carboidratos tem função ser estruturante e energético. Carboidratos estruturados por compostos orgânicos como carbono, oxigênio e hidrogênio.”

Após obter o texto, realizamos um processo de normalização usando técnicas de processamento de linguagem natural a fim de torná-lo adequado ao processamento por modelos computacionais. Cada texto é submetido a um conjunto de transformações de filtragem, sumarização, tokenização, formação de bigramas e trigramas, remoção de *stop*

words, lematização e substituição de conceitos semelhantes. Essas transformações são executadas em ordem estrita e sua inversão altera negativamente a qualidade dos resultados.

3.2.COMPARANDO DIFERENTES MAPAS CONCEITUAIS

Quando se trabalha com conteúdo de texto, uma preocupação é poder verificar semelhanças que vão além da comparação do conjunto de palavras e seja sensível a variações morfológicas e estruturais, uma vez que, quando alguma palavra do texto ou sua ordem mudam, isso pode também alterar seu significado. As sentenças “tenho um grande amigo” e “tenho um amigo grande”, por exemplo, apesar de possuírem as mesmas palavras, possuem diferentes semânticas. Por outro lado, as sentenças “estou indo pra minha casa” e “vou pro meu lar”, embora sejam diferentes morfológicamente, expressam o mesmo conteúdo semântico.

Para considerar esse fator e alcançar uma dimensão semântica, foi utilizado o modelo computacional Word2Vec (MIKOLOV, CHEN, CORRADO, & DEAN, 2013) de processamento de linguagem natural para medir a similaridade de palavras em sentenças. Ele permite comparar se duas palavras diferentes possuem o mesmo significado. Para isso, representa palavras como vetores de N-dimensões que consideram o contexto e permitem capturar semântica. A análise de similaridade semântica entre duas palavras é feita pela comparação de menor distância entre esses vetores por uma rede neural rasa pré-treinada. Dessa forma, vetores próximos no espaço vetorial têm significados semelhantes, com base no contexto, e vetores de palavras distantes têm significados diferentes. A suposição subjacente é que duas palavras que compartilham contextos semelhantes também compartilham um significado semelhante e, conseqüentemente, uma representação vetorial semelhante do modelo. Dessa forma, a verificação de similaridade é feita comparando os vetores, obtidos pela rede neural, para cada palavra em sentenças permitindo a verificação de similaridade entre palavras morfológicamente diferentes.

Para calcular o vetor do texto como um todo, foi utilizado o algoritmo Doc2Vec (LE & MIKOLOV, 2014) que gera um vetor considerando o relacionamento entre as palavras nele presentes. No Doc2Vec, os vetores de um documento são obtidos a partir da inferência realizada por uma rede neural. Assim, a rede é treinada na tarefa de prever uma palavra central com base na média dos vetores de palavras de contexto e do vetor de documentos completo do *corpus* (conjunto de documentos de texto), dando ao resultado uma dimensão semântica.

Para treinar e testar a rede neural que faz inferência de vetores, foram testadas duas abordagens de treinamento. Na primeira abordagem, o modelo foi treinado a partir de um *corpus* mais amplo, denominado text8, composto por 5 GB de artigos da Wikipédia, contando com cerca de 17 milhões de palavras sobre diferentes temas. A ideia principal era que o modelo pudesse ter conhecimento sobre diferentes assuntos e, assim, ter um escopo mais abrangente na comparação de possíveis assuntos em atividades propostas. Quando o modelo treinado foi avaliado, verificou-se que, embora apresentasse resultados positivos, tinha algumas variações para inferência de novos vetores, influenciando na comparação e formação de agrupamentos semânticos. Esse resultado ocorre quando o *corpus* de treinamento é muito diferente do *corpus* de teste, ou quando não possui amostras suficientes dele. Na segunda abordagem, foi utilizado para treinamento e teste, um *corpus* de treinamento privado, originado de mapas conceituais de diferentes atividades sobre o mesmo tema. Isso proporcionou excelentes resultados para inferência de novos vetores, apresentando uma consistência de validação cruzada do *corpus* frente ao aprendizado do modelo para a rede neural. Essa segunda abordagem proporcionou melhores resultados comparativos porque os documentos utilizados pertenciam ao mesmo tema, reduzindo as variações ocorridas na primeira abordagem devido ao escopo mais amplo e permitindo a construção e formação de um *corpus* temático específico, que é ampliado à medida que novos experimentos são realizados.

Depois que o modelo foi treinado e os vetores para os documentos de texto dos mapas conceituais foram inferidos, realizou-se comparações de similaridade usando duas medidas: a similaridade do cosseno (GAN, MA, & WU, 2007) e o cosseno suave de similaridade (SIDOROV, GELBUKH, GOMEZ-ADORNO, & PINTO, 2014). A similaridade do cosseno é uma medida de similaridade entre dois vetores que compara o ângulo entre os vetores normalizados resultantes. A medida do cosseno suave é um método que permite avaliar a similaridade entre dois documentos de forma significativa, mesmo quando eles não têm palavras em comum. Ela usa uma medida de semelhança entre os vetores que pode ser derivada da correspondência usando o Doc2Vec, fornecendo melhores resultados de semelhança semântica (CHARLET & DAMNATI, 2017).

3.3.AGRUPANDO ESTUDANTES A PARTIR DOS MAPAS CONCEITUAIS

Para realizar agrupamentos semânticos a partir de vetores inferidos dos textos, foram testados dois algoritmos para compor o *framework*: K-Means e BIRCH (GAN, MA, & WU, 2007). Inicialmente, escolheu-se o K-Means, no entanto, seu cálculo inicial do centroide é escolhido aleatoriamente, apresentando diferentes resultados para cada execução, sendo necessário executar o algoritmo várias vezes e escolher o resultado que produz os melhores grupos. O agrupamento BIRCH tem a vantagem de apresentar sempre o mesmo resultado, pois é gerado a partir da mesma função de distribuição em um processo de agrupamento hierárquico, podendo formar grupos de forma incremental e dinâmica. Por esta razão, o agrupamento BIRCH foi escolhido em vez de K-Means.

Para avaliar a qualidade do agrupamento gerado, utilizou-se duas medidas: a análise do valor de silhueta e análise do gráfico do cotovelo (SAPUTRA, SAPUTRA, & OSWARI, 2019). A análise de silhueta fornece a quantidade ideal de grupos, a partir da medição da distância de separação entre os grupos resultantes, realizando uma medida de quão próximo cada ponto em um grupo está de pontos em grupos vizinhos. A análise do gráfico do cotovelo considera a medida da soma dos quadrados das distâncias dos participantes do grupo ao centro do grupo mais próximo. Ambas as medidas são formas de ajudar a identificar o melhor número de agrupamentos sob diferentes abordagens.

Os métodos apresentados até aqui fornecem uma visão externa dos agrupamentos a partir de métricas que comparam a semelhança de indivíduos com intuito de atribuí-los a um grupo e a distância entre diferentes grupos. Complementarmente, desenvolvemos uma técnica (BOGUSKI & CURY, 2017) que permite a análise interna de grupos a partir do conhecimento de como eles estão estruturados e associados em termos de conceitos e relacionamentos que os constituem. Essa perspectiva é proporcionada com uso de regras de associação e exprime como os conceitos são percebidos e representados coletivamente dentro do grupo, fornecendo um conhecimento mais profundo de como o indivíduo estrutura e relaciona conceitos na construção de mapas. Essa perspectiva é complementar às anteriores e contribui substancialmente na análise de agrupamentos.

4. O FRAMEWORK

O *framework* reúne, de modo estruturado, um conjunto de métodos, técnicas e estratégias desenvolvidos a fim de permitir a construção e análise de agrupamentos temáticos e semânticos sobre mapas conceituais. Ele é fundamentado em 5 pilares, conforme Figura 2.

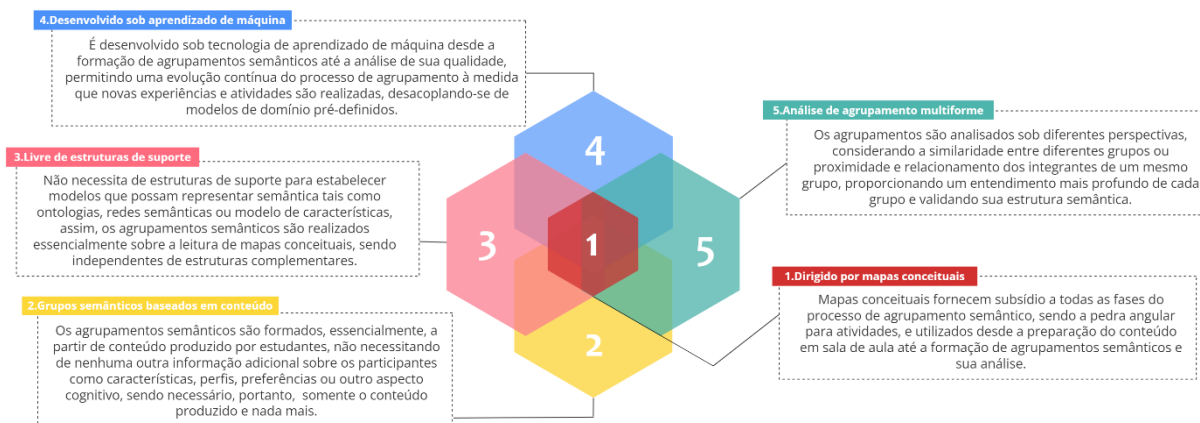


Figura 2. Os pilares do framework

O modelo conceitual do *framework* proposto é constituído de 5 fases: preparação, aplicação, processamento, agrupamento temático, agrupamento semântico, análise associativa de agrupamento (Figura 3).

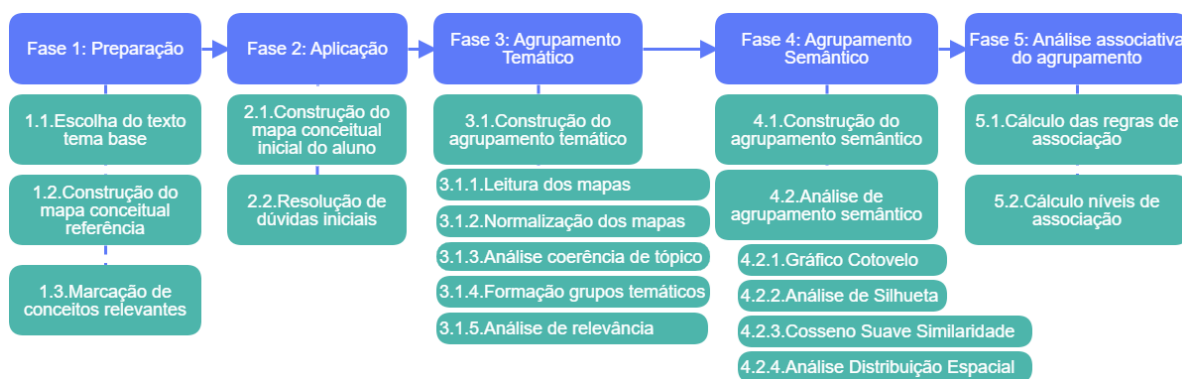


Figura 3. Modelo conceitual do framework

Na fase de preparação o professor planeja as atividades de ensino-aprendizagem, definindo o tema, conteúdo, conceitos relevantes e um mapa conceitual de referência. É um estágio crucial para garantir que os mapas conceituais produzidos pelos alunos estejam alinhados com os objetivos educacionais.

Na fase de aplicação, os alunos produzem mapas conceituais baseados nas atividades propostas. O papel do professor é de suporte, sem interferir na construção dos mapas, garantindo a autenticidade do trabalho dos alunos.

A fase de agrupamento temático contempla estratégias para agrupar estudantes sob a perspectiva temática. Em (BOGUSKI & CURY, 2022), aplicamos um conjunto de técnicas para formar grupos temáticos. A comparação temática é um passo anterior à comparação semântica, com granularidade maior e menor complexidade, permitindo diferenciar se duas informações pertencem ao mesmo assunto, pois, se não pertencerem, por mais vigorosos que sejam os esforços comparativos, não haverá acordo, tampouco aproximação semântica, dentro de diferentes temas. Embora não permita verificar se os participantes estão dizendo a mesma coisa, pode indicar, por outro lado, que falam da mesma coisa, e isso é relevante, por exemplo, para a atuação de um mediador na resolução de divergências entre eles.

Na fase de agrupamento semântico é realizado o agrupamento e análise semântica dos mapas conceituais usando uma rede neural treinada para reconhecer similaridade semântica. Avalia-se a formação dos grupos semânticos, comparando-os com os grupos temáticos identificados na fase anterior. São utilizadas métricas como o método do cotovelo, o método silhueta e a similaridade do cosseno suave para avaliar a qualidade dos agrupamentos semânticos.

A análise associativa de agrupamento permite examinar a estrutura interna dos grupos formados, utilizando regras de associação para entender como os conceitos dentro de cada agrupamento estão associados entre si. Essa análise oferece uma visão mais detalhada das relações e estruturas dos conceitos nos mapas conceituais.

Essas cinco fases permitem uma análise abrangente e detalhada dos agrupamentos temáticos e semânticos, fornecendo compreensão valiosa sobre a construção do conhecimento dos alunos e a eficácia das atividades pedagógicas. Para que o *framework* pudesse ser instanciado, produzimos um ambiente de software que suportasse as cinco fases em atividades pedagógicas.

5. EXPERIMENTOS E RESULTADOS

Para validação, utilizamos experimentos em sala de aula e questionários respondidos pelos professores quanto ao uso e eficácia do *framework* para formar e analisar grupos semânticos. Antes de utilizarmos o *framework* em experimentos em sala de aula, realizamos

em torno de 500 testes em laboratório para seu refinamento e sintonia, a fim de assegurar a capacidade em formar grupos semânticos adequadamente.

Nos experimentos, comparamos os agrupamentos semânticos realizados por professores com os realizados pelo *framework*. No primeiro experimento, utilizamos os mapas conceituais produzidos em uma atividade de uma turma de 28 alunos do curso de arquivologia do departamento de Arquivologia da Universidade Federal do Espírito Santo – UFES, na disciplina de Organização e Representação do Conhecimento, no semestre 2018/1, sob a tutela da professora Tânia Barbosa Salles Gava. O objetivo da disciplina era capacitar o aluno na construção, organização e representação do seu próprio conhecimento, usando como apoio uma ferramenta tecnológica para construção de mapas conceituais. A atividade consistiu na livre construção de mapas conceituais tendo como base, unicamente, o conhecimento tácito que cada aluno possuía de uma lista fechada de conceitos apresentada pela professora. Assim, não foi utilizado material anexo que pudesse influenciar sua construção. O tema proposto foi “como a tecnologia da informação pode ser utilizada como estratégia de preservação digital de acervos digitais”. Apesar de a lista de conceitos ser fixa, a mudança no relacionamento entre os conceitos conduz a mapas com diferentes conteúdos semânticos e estávamos procurando verificar se o *framework* conseguia medir essas nuances.

Para a validação dos agrupamentos semânticos, neste primeiro experimento, pedimos que um grupo de 3 pessoas, que chamamos de especialistas, realizassem agrupamentos, a partir da inspeção dos mapas. Depois, foram executadas as fases do *framework* que conduzem aos agrupamentos no ambiente de software. O grupo de especialistas propôs 10 grupos e o *framework* 4 grupos. Essa divergência ocorreu pela diferenciação semântica excessiva entre os elementos dos grupos, por parte dos especialistas, a ponto de formar novos grupos. Verificou-se que um resultado semelhante é apresentado pelo *framework* quando consideramos o mesmo número de grupos semânticos propostos pelos especialistas. Contudo, é justamente para evitar esse tipo de situação que utilizamos o cálculo do valor de silhueta e gráfico de cotovelo, os quais nos permite obter o número de melhores agrupamentos. Os grupos semânticos propostos pelos especialistas foram $S_E = [\{A03, A06, A14, A16, A19, A25\}, \{A23\}, \{A01, A02\}, \{A04, A05, A07, A08, A10, A11, A15, A20, A22, A26\}, \{A27\}, \{A21, A28\}, \{A12, A13, A17\}, \{A18\}, \{A24\}, \{A09\}]$ e pelo *framework* foram $S_F = [\{A03, A06, A14, A16, A19, A25, A23\}, \{A01, A02, A04, A05, A07, A08, A10, A11,$

A15, A20, A22, A26, A27},{A21, A28},{A12, A13, A17, A18, A24, A09}], onde, A01 representa o aluno A01 e assim por diante.

Para a análise de agrupamentos no contexto proposto, escolhemos o coeficiente de Rand (RI) (WARRENS, 2022), dentre as métricas que consideramos relevantes para medir agrupamentos (ZHANG, 2021), (HUBERT, 2022), (FOWLKES & MALLOWS, 2022), (SMITH, SMITH, & ZHANG, 2021), (MATTHEWS, 2021), (XU, 2021), pois é simples para agrupamentos pequenos que possuem critérios bem definidos de formação, como os propostos, tendo aplicação direta, não requerendo ajustes complexos ou suposições adicionais sobre a distribuição dos dados.

A comparação com o coeficiente de Rand apontou uma similaridade de 87,30% para as melhores correspondências entre os agrupamentos. Após a análise de resultados propostos pela *framework*, os especialistas reviram os agrupamentos semânticos, ajustando a proposta inicial para a seguinte configuração $S_E = [\{A03, A06, A14, A16, A19, A25\}, \{A01, A02, A04, A05, A07, A08, A10, A11, A22, A26, A27, A23\}, \{A21, A28\}, \{A12, A13, A17\}, \{A09, A18, A24\}]$. Isso aumentou a correspondência para 94,4% entre os grupos semânticos.

O segundo experimento foi realizado em uma turma, pertencente à escola EEEFM Narceu de Paiva Filho, localizada no município de Ibiráçu, no estado de Espírito Santo. Participaram dessa atividade, 20 alunos, porém 1 dos mapas produzidos foi um mapa mental, impossibilitando assim seu uso no experimento. Os alunos cursaram a disciplina de “Morfologia Humana” na 3ª série do ensino médio do semestre 2024/1, ministrado pela professora Renata Carrara Torri. A disciplina de morfologia humana objetiva estudar como as diversas partes do corpo são responsáveis pelo desempenho esportivo. O tema escolhido foi “O que é testosterona”. Esse conteúdo já vinha sendo ministrado em sala de aula ao longo do período letivo com aulas expositivas, artigos, material para discussão.

A atividade do experimento consistiu na construção de mapas conceituais a partir de um texto base fornecido pela professora. A atividade foi realizada em dois dias. No primeiro dia a professora trabalhou o texto base com os alunos. No segundo dia, os alunos receberam novamente o texto base e a partir dele, foram orientados a construir seu mapa conceitual. A produção de mapas foi feita individualmente e, portanto, não houve interação entre os alunos. Solicitamos à professora que realizasse os agrupamentos semânticos que ela percebia a partir da análise dos mapas conceituais dos alunos. Os agrupamentos semânticos formados pela professora foram $S_P = [\{A06, A11, A17, A14\}, \{A19\}, \{A01, A15, A04, A07, A03,$

A16},{A08, A10, A02, A05, A09, A12, A13, A18}] e $S_F=[\{A06, A11, A17, A19\},\{A01, A15, A04, A07, A03, A16, A08, A10\},\{A02, A05, A09, A12, A13, A18, A14\}]$, pelo *framework*. O valor das melhores correspondências foi de 62,92%. Após a análise dos resultados, os novos grupos semânticos propostos pela professora foram $S_P=[\{A06, A11, A17, A19\},\{A01, A15, A04, A07, A03, A16\},\{A08, A10, A02, A05, A09, A12, A13, A14, A18\}]$, aumentando a correspondência para 91,30%.

O terceiro experimento também foi realizado na escola EEEFM Narceu de Paiva Filho, numa turma da 2ª série do ensino médio do semestre 2024/1, constituída por 25 alunos, com idades entre 15 e 17 anos, cursando a disciplina de “Química e Esportes” sob a tutela da professora Adriany Devens. Dos 25 alunos matriculados regularmente na disciplina, 19 pessoas participaram deste experimento. A atividade consistiu na livre construção de mapas conceituais a partir, exclusivamente, do conhecimento que tinham sobre o tema, quer seja por ensino ministrado em sala de aula por meio de artigos, material para discussão ou experiências cotidianas dentro e fora da escola. O tema escolhido foi “o uso de química nos esportes”, aprofundando-se principalmente no aprendizado de Química Orgânica. O conhecimento dos alunos sobre a construção de mapas conceituais era iniciante, desse modo, a produção de mapas conceituais foi também um novo conhecimento a ser assimilado. Para os agrupamentos semânticos, a média das similaridades dos grupos foi 57,60%. Os grupos propostos pela professora e *framework* foram, respectivamente $S_P=[\{A02, A03, A04, A07, A08, A11, A14, A19, A16\},\{A17, A18\},\{A01, A05, A06\},\{A09, A12, A15\},\{A10, A13\}]$ e $S_F=[\{A02, A03, A04, A07, A08, A17\},\{A01, A05, A06, A12, A14, A15, A11, A18, A19\},\{A09, A10, A13, A16\}]$. Após a análise dos resultados, os novos grupos propostos pela professora foram $S_P=[\{A02, A03, A04, A07, A08, A17, A19\},\{A01, A05, A06, A12, A14, A15, A11, A13, A18\},\{A09, A10, A16\}]$, ajustando o valor de semelhança para 83,06%.

A seguir, apresentamos, na Figura 4, Figura 5 e Figura 6, a disposição espacial e gráfico de cotovelo dos agrupamentos semânticos para os experimentos 1, 2 e 3, respectivamente, realizados pelo *framework*. Essas perspectivas são diferentes, mas complementares. Enquanto o gráfico de cotovelo indica o número ideal de grupos para o experimento, a disposição espacial apresenta a distribuição dos elementos em cada grupo, permitindo que os professores tenham uma visão mais abrangente da proximidade dos

grupos e dos elementos de cada grupo. Dessa forma, grupos mais densos denotam maior uniformidade semântica enquanto que grupos dispersos, apesar de possuírem semântica em comum, necessitam ser melhor trabalhados.

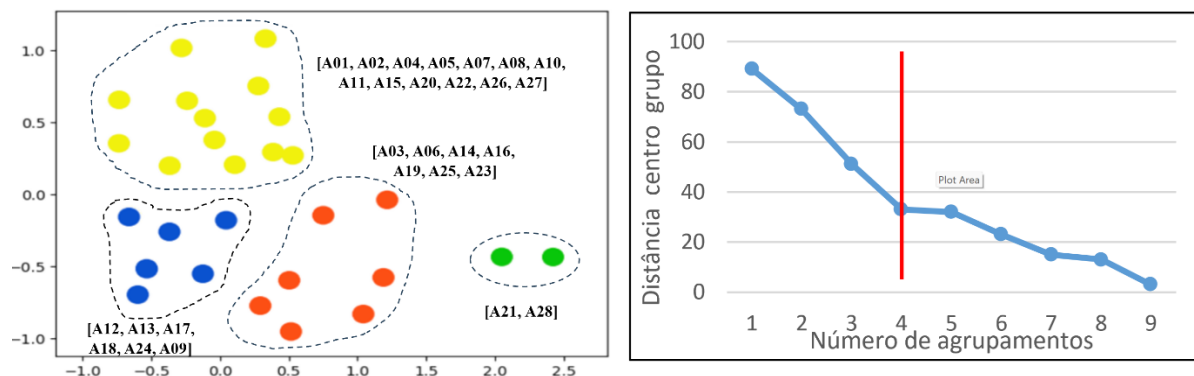


Figura 4 – Disposição dos agrupamentos e gráfico de cotovelo para o experimento 1

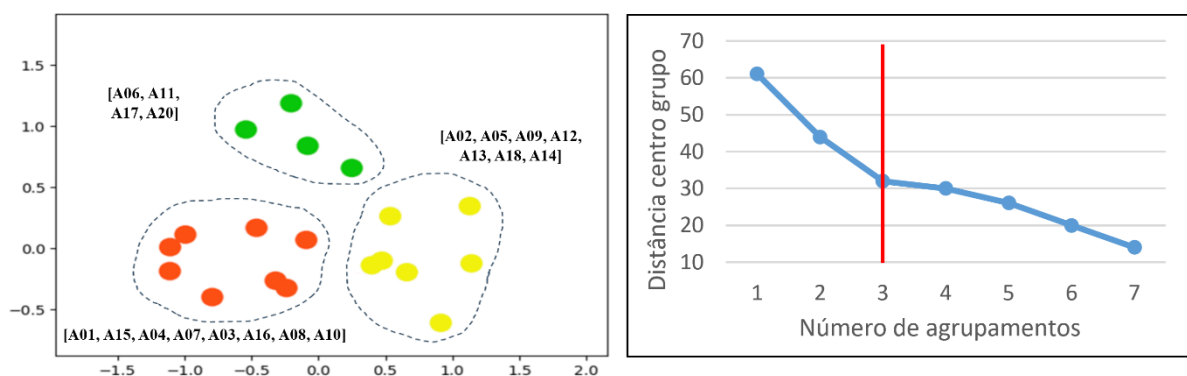


Figura 5 – Disposição dos agrupamentos e gráfico de cotovelo para o experimento 2

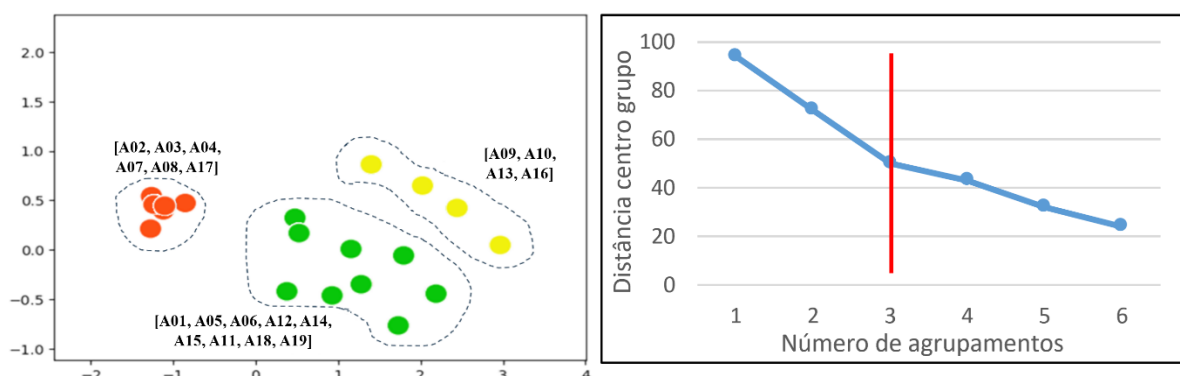


Figura 6 – Disposição dos agrupamentos e gráfico de cotovelo para o experimento 3

Para outras análises de agrupamentos e similaridade semântica, fornecemos duas interfaces. A primeira, apresenta uma tabela comparativa de proximidade semântica de um mapa em relação aos demais. Para validá-la, selecionamos um mapa e pedimos que a

professora classificasse, em ordem decrescente de similaridade, quais mapas julgava mais semelhantes a ele. Os resultados para esse teste obtiveram uma assertividade de 87,8%.

A segunda, apresenta como os conceitos estão associados nos mapas de cada grupo. Sua validação consistiu em comparar as regras de associação com os conceitos relevantes marcados pelas professoras. Neste caso, é esperado que os conceitos relevantes devam aparecer em maior frequência e associação nos mapas dos alunos.

6. CONCLUSÕES

Neste trabalho, apresentamos um *framework* para a formação e análise de agrupamentos semânticos de alunos a partir da leitura automatizada dos mapas conceituais por eles produzidos. O *framework* é estruturado de modo que os agrupamentos sejam realizados apenas a partir dos mapas, não dependendo de estruturas de suporte ou modelos que os representem previamente. Os experimentos em sala de aula permitiram a validação em diferentes cenários.

No primeiro experimento, o objetivo era medir a similaridade semântica de mapas gerados a partir de uma mesma lista de palavras, porém, reordenadas de modo a formar diferentes sentenças nos mapas de cada aluno. No segundo experimento, a abordagem utilizada preocupou-se em validar se o *framework* pode formar, adequadamente, grupos semânticos em situações em que mapas foram produzidos a partir de uma mesma referência, como um texto base, objetivando verificar se era possível capturar nuances contextuais a partir de um mesmo conteúdo base, uma vez que é esperado que mapas construídos a partir de um mesmo texto base sejam mais uniformes. No terceiro experimento, a preocupação era validar a formação de grupos a partir dos mapas construídos, exclusivamente, pelo conhecimento dos alunos sobre um assunto, sem a utilização de material anexo que pudesse influenciar sua construção.

Deste modo, esperava-se medir similaridade semântica em mapas com maior diversidade de conteúdo e contexto. Nos 3 casos, pudemos validar a formação de grupos em cenários estocásticos decrescentes, obtendo resultados de 94,4%, 91,30% e 83,06%, respectivamente, de similaridade entre os grupos formados pelas professoras e pelo *framework* em cada experimento. Além disso, o *framework* mostrou-se útil e eficaz na

formação e análise de agrupamentos semânticos a ponto de permitir que as professoras ajustassem o agrupamento realizado inicialmente.

Após os experimentos, submetemos um formulário de avaliação, em escala likert, às professoras com 25 questões relacionadas a diferentes aspectos do *framework* nas categorias “realização da atividade”, “formação de agrupamentos temáticos”, “formação de agrupamentos semânticos”, “análise de agrupamentos”, “apresentação e interpretação de resultados”, “utilidade para uso em atividades em sala de aula” obtendo “Muito Bom”, “Excelente”, “Excelente”, “Muito Bom”, “Muito Bom” e “Excelente”, respectivamente. Reiteramos que os resultados propostos pelo *framework*, embora sejam bastante assertivos, tem o propósito de auxiliar o professor no discernimento e escolha de melhores agrupamentos, devendo ser vistos mais como diagnósticos que prescritivos.

Agrupamentos semânticos podem ser ferramentas poderosas para professores em atividades pedagógicas, auxiliando de maneira significativa na personalização do ensino, organização de conteúdo, facilitação da aprendizagem colaborativa, engajamento dos alunos e avaliação mais eficaz. Ao agrupar estudantes com base em semelhanças de compreensão de determinados conceitos, os professores conseguem identificar lacunas no conhecimento e adaptar o ensino de acordo com as necessidades específicas de cada grupo. Isso permite propor desafios adequados ao nível de entendimento, promovendo um aprendizado contínuo e evitando sobrecargas ou subestimações.

Além disso, agrupamentos semânticos facilitam a troca de conhecimento e a resolução de problemas colaborativamente, já que os alunos compartilham um entendimento comum dos temas abordados, resultando em discussões mais produtivas e em um aprendizado mais rico. Também aumentam o engajamento dos alunos ao reunir indivíduos com interesses semelhantes, proporcionando um ambiente de aprendizado mais motivador e seguro, especialmente ao reduzir a ansiedade de alunos que se sentem pressionados em contextos competitivos. Para os professores, esses agrupamentos tornam a avaliação mais direcionada, permitindo monitorar o progresso coletivo e individual de maneira mais precisa, enquanto o *feedback* pode ser mais eficaz e relevante para o grupo como um todo.

O uso do *framework* apresentado em atividade de ensino-aprendizagem pode otimizar ainda mais esse processo, ajudando a identificar padrões nos conhecimentos dos alunos e a organizar os grupos de forma mais eficiente. Com isso, os professores podem não só economizar tempo na gestão de diferentes níveis de compreensão em sala de aula, como também promover o desenvolvimento do pensamento crítico, incentivando discussões mais

aprofundadas e complexas dentro dos agrupamentos, onde os estudantes se sentem confortáveis para explorar temas avançados em um ambiente de suporte mútuo. Dessa forma, os agrupamentos semânticos oferecem uma abordagem estruturada e adaptativa para maximizar o aprendizado, permitindo uma educação mais inclusiva e eficaz em diversos contextos pedagógicos, seja em ambientes presenciais ou digitais.

Como continuidade deste trabalho, objetivamos considerar nos agrupamentos, as implicações significantes (PIAGET & GARCIA, 1989), (DUTRA, 2018) sobre mapas conceituais a fim detectar níveis de entendimento dos alunos sobre o assunto. Finalmente, enfatizamos que a verificação de semântica sobre mapas conceituais, com uso de processamento de linguagem natural é um campo relativamente emergente, permitindo o desenvolvimento de uma série de estudos e pesquisas capazes de gerar uma grande contribuição para os processos de ensino e aprendizagem.

REFERÊNCIAS

- BODENREIDER, O., & MCCRAY, A. T. (2003). Exploring semantic groups through visual approaches. *Journal of Biomedical Informatics*, 414-432. doi:https://doi.org/10.1016/j.jbi.2003.11.002
- BOGUSKI, R. R., & CURY, D. (2017). Usando regras de associação para a identificação de falhas conceituais em estudantes. *XXIX Simpósio Brasileiro de Informática na Educação (Brazilian Symposium on Computers in Education)*, (pp. 1443-1452).
- BOGUSKI, R. R., & CURY, D. (2022). Agrupando estudantes a partir da similaridade semântica de mapas conceituais. *2022: ANAIS DO XXXIII SIMPÓSIO BRASILEIRO DE INFORMÁTICA NA EDUCAÇÃO*, (pp. 979-991). doi:https://doi.org/10.5753/sbie.2022.224957
- CHARLET, D., & DAMNATI, G. (2017). Soft-Cosine Semantic Similarity between Questions for Community Question Answering. *Proceedings of the 11th International Workshop on Semantic Evaluation*, (pp. 315–319). Vancouver, Canada.
- CIUBUC, C., DASCALU, M., TRAUSAN-MATU, S., & MARHAN, A. -M. (2013). Forming Teams by Psychological Traits -- An Effective Method of Developing Groups in an Educational Environment. *2013 19th International Conference on Control Systems and Computer Science*, (pp. 597-602). Bucharest, Romania. doi:10.1109/CSCS.2013.54.

DRAGOMIR, D., D., S. B., & M., A.-M. M. (2020). Assessing the Perception of Abstract Paintings with Elements of Augmented Reality. *Journal of Augmented and Virtual Reality*, (pp. 54-67). doi:10.1007/s40940-020-00048-3

DUTRA, Í. M. (2018). Obtido em 16 de 09 de 2024, de https://peadalvorada5.pbworks.com/f/italo_mapas.pdf

GAN, G., MA, C., & WU, J. (2007). *Data Clustering: Theory, Algorithms, and Applications*. SIAM.

HAN, J., KAMBER, M., & PEI, J. (2011). *Data Mining Concepts and Techniques*. Elsevier.

JACOB DEVLIN, M.-W. C. (2018). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT 2018)*, (pp. 4171-4186). New Orleans, LA. doi:10.48550/arXiv.1810.04805

LAMAS, F., BOERES, C., CURY, D., & MENEZES, C. S. (2005). Comparando mapas conceituais utilizando correspondência de grafos. *Simpósio Brasileiro De Informática Na Educação - SBIE*, (pp. 24-27).

LE, Q. V., & MIKOLOV, T. (2014). Distributed Representations of Sentences and Documents. *Proceedings of the 31st International Conference on International Conference on Machine Learning*, (pp. 1188-1196).

LIMONGELLI, C., SCIARRONE, F., LOMBARDI, M., MARANI, A., & TEMPERINI, M. (2017). A framework for comparing concept maps. *16th International Conference on Information Technology Based Higher Education and Training (ITHET)*.

LOVATI, M. P., CURY, D., AGUIAR, C. Z., & AZEREDO, R. (2017). Clusterizando mapas conceituais para identificar desempenho cognitivo em grupos. *XXVIII Simpósio Brasileiro de Informática na Educação - SBIE (Brazilian Symposium on Computers in Education)*. doi:10.5753/cbie.sbie.2017.1397

MANSKE, S., & HOPPE, H. U. (2017). Managing Knowledge Diversity: Towards Automatic Semantic Group Formation. *IEEE 17th International Conference on Advanced Learning Technologies (ICALT)*, (pp. 330-332). doi:10.1109/ICALT.2017.132.

MIKOLOV, T., CHEN, K., CORRADO, G., & DEAN, J. (2013). Efficient Estimation of Word Representations in Vector Space. *ICLR Workshop Papers*.

MIKOLOV, T., CHEN, K., CORRADO, G., & DEAN, J. (2013). Efficient Estimation of Word Representations in Vector Space. *ICLR Workshop Papers*.

MOREIRA, M., & RODRÍGUEZ, M. C. (1997). Aprendizagem Significativa: Um Conceito Subjacente. *Actas del Encuentro Internacional sobre el Aprendizaje Significativo*, (pp. 19-44). Burgos.

MOUBAYED, A., INJADAT, M., SHAMI, A., & LUTFIYYA, H. (2020). Student Engagement Level in an e-Learning Environment: Clustering Using K-means. *American Journal of Distance Education*, 137-156.

MUHAMMAD, F., & NORDIN, A. (2013). Forming Community of Practice in IT department of Higher Education Institution. **2013 International Conference on Research and Innovation in Information Systems (ICRIIS)**, (pp. 67-72). Kuala Lumpur, Malaysia. doi:10.1109/ICRIIS.2013.6716687

NOVAK, J., & GOWIN, D. (1984). **Learning how to learn**. Cambridge: Cambridge University Press.

OTHMANE, Z. B., & ALI, A. H. (2016). Group Formation with Incomplete Profiles. **12th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS)**, (pp. 732-745). Naples, Italy. doi:10.1109/SITIS.2016.122

OUNNAS, A., DAVIS, H., & MILLARD, D. (2008). A Framework for Semantic Group Formation. **2008 Eighth IEEE International Conference on Advanced Learning Technologies**, (pp. 34-38). Santander, Spain. doi:10.1109/ICALT.2008.226.

PETERS, S., & SHROBE, H. E. (2003). Using Semantic Networks for Knowledge Representation in an Intelligent Environment. **Proceedings of the First IEEE International Conference on Pervasive Computing and Communications, 2003**, (pp. 323-329).

PIAGET, J., & GARCIA, R. (1989). *Hacia una Lógica de Significaciones*. Gedisa, México.

SANTOS, A. M. (2009). The AMiGO-Mob: Agent-based Middleware for Group-oriented Mobile Service Provisioning. **International Journal of Computer Applications in Technology**, (pp. 262-272). doi:10.1504/IJCAT.2009.028289

SAPUTRA, D. M., SAPUTRA, D., & OSWARI, L. D. (2019). Effect of Distance Metrics in Determining K-Value in K-Means Clustering Using Elbow and Silhouette Method. **Sriwijaya International Conference on Information Technology and Its Applications**.

SHAVLIK, J., & MCNELIS, J. A. (2012). Visual Exploration of Relationships and Structure in Low-Dimensional Embeddings. **Proceedings of the 2012 International Conference on Data Mining**, (pp. 32-40). Washington, DC. doi:10.1109/ICDM.2012.38

SIDOROV, G., GELBUKH, A., GOMEZ-ADORNO, H., & PINTO, D. (2014). Soft Similarity and Soft Cosine Measure: Similarity of Features in Vector Space Model. **Computación y Sistemas**, 491–504.

SIMÓN, A., L. C., & ROSETE, A. (2007). Generation of OWL Ontologies from Concept Maps in Shallow Domains. **Congresos de la Asociación Española para la Inteligencia Artificial (CAEPIA)**, (pp. 259-267).

THANGARAJ, M., & MEENATCHI, V. T. (2015). Applying prefetching in Online Social Network to gain social intelligence. **2015 International Conference on Computer Communication and Informatics (ICCCI)**, (pp. 1-4). Coimbatore, India. doi:10.1109/ICCCI.2015.7218066.

VINELLA, F. L., LYKOURTZOU, I., & MASTHOFF, J. (2021). Users' Preferences of Profiling Attributes on Crowdsourcing Team Formation Systems. **021 16th International**

Workshop on Semantic and Social Media Adaptation & Personalization (SMAP), (pp. 1-10). Corfu, Greece. doi:10.1109/SMAP53521.2021.9610773.

ZHUGE, H. (2009). Communities and Emerging Semantics in Semantic Link Network: Discovery and Learning. ***IEEE Transactions on Knowledge and Data Engineering***, 785-799. doi:10.1109/TKDE.2008.141