

## Organização e representação de conhecimento de temas de pesquisa

Eduardo Amadeu Dutra Moresi<sup>1,3</sup>, Ivo Pierozzi Júnior<sup>2</sup>, Leandro Henrique Mendonça de Oliveira<sup>2</sup>, Alessandra de Moura Brandão<sup>3</sup>

moresi@ucb.br; emoresi@cgee.org.br; ivo.pierozzi@embrapa.br; leandro.oliveira@embrapa.br; abrandao@cgee.org.br

<sup>1</sup> Universidade Católica de Brasília, QS 07 Lote 01 EPCT Bloco N, 71966-700, Brasília, DF, Brasil

<sup>2</sup> Afiliação do Autor Dois, Morada, Código postal, Localidade, País

<sup>3</sup> Centro de Gestão e Estudos Estratégicos, Morada, Código postal, Localidade, País

DOI: 10.17013/risti.33. 63–77

**Resumo:** Este artigo apresenta e descreve uma metodologia de organização e representação do conhecimento de temas de pesquisa, cuja execução permite o alinhamento de várias ferramentas objetivando sistemas de organização de conhecimento como recurso para facilitar a elaboração de sistemas de conceitos, expressões de busca e significação da informação recuperada em bases de dados. A metodologia proposta se constitui de quatro etapas, baseadas em métodos e técnicas de: mapeamento do domínio do conhecimento; codificação do conhecimento; aplicação de linguística de *corpus* e de processamento de linguagem natural e representação do conhecimento. Um exemplo prático é apresentado sobre o domínio temático da inteligência artificial na educação, a partir da recuperação dos metadados de uma pesquisa bibliográfica na base Scopus. Em seguida, é realizada a análise da rede de coocorrência de palavras-chave, que revela o espaço conceitual que emerge da literatura recuperada.

**Palavras-chave:** mapeamento do conhecimento; organização do conhecimento; representação do conhecimento; espaços conceituais; inteligência artificial na educação.

### ***Knowledge organization and representation: methodological and technological increments for conceptual mapping***

**Abstract:** This article presents and describes a methodology of knowledge organizing and representing of research themes, whose execution allows the alignment of several tools aiming, systems of knowledge organization as a resource to facilitate the elaboration of concept systems, search expressions and information signification recovered from databases. The proposed methodology consists of four steps based on methods and techniques of: mapping the knowledge domain; knowledge codification; linguistic corpus applications and natural language processing; and knowledge representation. A practical example is presented about the thematic domain of artificial intelligence in education, based on the retrieval

of bibliographic metadata from the Scopus database. Then, the analysis of the keyword co-occurrence network is performed, revealing the conceptual space that emerges from the retrieved literature.

**Keywords:** knowledge mapping; knowledge organization; knowledge representation; conceptual spaces; artificial intelligence in education.

## 1. Introdução

Na era do conhecimento, os processos de pesquisa, desenvolvimento e inovação, devido ao alto grau de complexidade e multi, inter e transdisciplinaridade temática, têm sido executados, cada vez mais, por meio das chamadas redes de conhecimento (Torres, Pierozzi, Pereira & Castro, 2011). Boa parte do conhecimento gerado nessas redes costuma ser explicitado em linguagem natural e, via de regra, textualmente, constituindo um volume enorme de documentos impressos ou digitais. Neste último formato, os documentos são disponibilizados amplamente em ambiente Web ou bases de dados, cujo acesso é facilitado pelas tecnologias da informação e comunicação (TIC).

Os mapas semânticos, construídos por meio de análises de coocorrência de palavras, são usados para entender a estrutura cognitiva de um domínio do conhecimento (Cahlik, 2000; Salvador & Lopez-Martinez, 2000). Esses mapas são gerados a partir de diferentes fontes textuais, incluindo palavras extraídas de títulos de artigos, palavras-chave declaradas pelos autores ou descritores atribuídos pelo editor, que estão registrados e indexados em bases de dados como a Elsevier Scopus e a Clarivate Web of Science. Atualmente, essas bases oferecem excelentes recursos bibliométricos para os usuários interessados em exploração e reuso de dados, informações e conhecimento.

O poder computacional para recuperação e análise desses conteúdos está prontamente disponível para os usuários analisarem redes com grandes quantidades de documentos (Zhao & Strotmann, 2008). Por outro lado, grande quantidade de informação requer o uso de ferramentas avançadas para viabilizar sua análise (Shiffrin & Börner, 2004), organização e sua ressignificação em outros contextos.

Historicamente, essas soluções não surgiram sem grande esforço. Dada a falta de recursos de computação, os primeiros estudos, naturalmente, tendiam a se concentrar em pequenas iniciativas e foram, com algumas exceções, acadêmicos por natureza. No entanto, o esforço para analisar grandes quantidades de documentos e obter respostas diminuiu consideravelmente, acompanhando o aumento da disponibilidade de dados em formato digital e o aumento exponencial do poder computacional, que incluem algoritmos avançados para visualização da informação e análises qualitativas.

Essas práticas configuram as chamadas competências analíticas (Levine, Strother-Garcia, Hirsh-Pasek, & Golinkoff, 2017) de uma organização e representam o conjunto de metodologias e ferramentas para reunião e análise de grandes volumes de informação com objetivo de criação de inteligência<sup>1</sup>. Atualmente, as competências analíticas resgatam um referencial teórico e metodológico multidisciplinar, que envolve, entre as principais abordagens, temas como visualização da informação, sistemas de organização

---

<sup>1</sup> “Inteligência é a informação avaliada” (Clauser & Weir, 1976)

do conhecimento (SOC), mineração de dados e de textos, processamento em linguagem natural (PLN), linguística de corpus e terminologias, mapas e estruturas conceituais, além de modelagens matemática e estatística.

O objetivo deste artigo é apresentar e descrever uma metodologia de mapeamento de domínios de conhecimento, cuja execução permite o alinhamento de várias ferramentas tendo-se como resultado, sistemas de organização de conhecimento como recursos para facilitar a elaboração de sistemas de conceitos, expressões de busca e significação da informação recuperada. O texto está estruturado, após esta breve introdução, com o referencial teórico, que aborda os principais conceitos que sustentam o trabalho, com a metodologia proposta, um exemplo prático, conclusões e referências bibliográficas.

## 2. Referencial Teórico

Sistemas de organização do conhecimento (SOC) se referem a um termo genérico usado para abordar uma ampla gama de itens (taxonomias, tesouros, esquemas de classificação e ontologias), que foram concebidos com diferentes propósitos, em momentos históricos distintos. Os SOC são caracterizados por diferentes estruturas e funções específicas, variados modos de se relacionar com a tecnologia e empregados em uma pluralidade de contextos por diversas comunidades. No entanto, o que todos eles têm em comum é o fato de terem sido projetados para apoiar a organização do conhecimento e da informação visando facilitar o gerenciamento e a recuperação. Para torná-lo acessível e utilizável (por agentes humanos ou tecnológicos), o conhecimento precisa, de fato, ser organizado de alguma maneira (Soergel, 2008), algo que, dada a grande quantidade de produção científica e cultural, tornou-se cada vez mais importante nos últimos anos.

A visualização de domínios do conhecimento tem como objetivo revelar as áreas da comunicação científica, conforme refletidos na literatura científica publicada e nas trajetórias de citação realizadas por pesquisadores em suas publicações. Há de fato uma conexão importante entre a visualização de domínios e o que Hjørland (1997) chamou de análise de domínio. A visualização fornece técnicas de habilitação para análise de domínio, especialmente em áreas de conhecimento multidisciplinares e de rápida evolução. O campo de visualização de domínio também tem sido chamado de cientografia (Garfield, 1994), embora este termo não pareça ser amplamente utilizado.

Os avanços na visualização da informação têm sido significativamente impulsionados pela pesquisa de recuperação de informação, cujo problema central tem sido como melhorar a sua eficiência e a sua eficácia. De um modo geral, quanto mais um usuário souber sobre seu espaço de pesquisa, maior será a probabilidade de sua pesquisa ser eficaz. Muitos sistemas de visualização de informações descrevem a estrutura semântica geral de uma coleção de documentos. Os usuários podem usar essa visualização estrutural como base para navegação e pesquisa subsequentes.

A seleção de uma unidade de análise é o passo inicial para explorar qualquer domínio do conhecimento, seguido pelas perguntas a serem respondidas. Essa premissa do processo vai ao encontro do espaço conceitual de Gärdenfors (2014). Este autor forneceu caracterizações precisas de várias categorias ontológicas tais como: um objeto é representado por uma sequência de pontos em um conjunto de espaços conceituais;

uma propriedade por uma região em um espaço conceitual; um conceito por uma sequência de regiões em um conjunto de espaços conceituais.

As unidades de análise mais comuns no mapeamento de literatura são periódicos, documentos, autores e termos descritivos ou palavras-chave. Cada um apresenta diferentes facetas de um domínio e facilita diversos tipos de análise. Por exemplo, um mapa de periódicos pode ser usado para obter uma visão macro da ciência (Bassecoulard & Zitt, 1999), mostrando as posições relativas e as relações entre as principais disciplinas.

Em muitas situações, as pessoas utilizam palavras com a finalidade de pesquisar assuntos, conceitos ou ideias. A intenção é recuperar informações que possibilitem o entendimento conceitual a partir do significado das relações que se estabelecem com os termos recuperados na pesquisa. A análise dos dados oriundos de metadados dos documentos indexados em uma base de referências bibliográficas torna-se essencial para gerenciar, localizar, interpretar ou usar outros dados ou uma fonte desses dados. Os metadados de uma referência bibliográfica não residem no objeto de informação, mas no seu relacionamento com outro objeto informacional e, mais especificamente, no seu uso. Nesse sentido, a organização do conhecimento se destina a atender as seguintes necessidades (Soergel, 2008): apoiar o pensamento, a construção de sentido, a integração do conhecimento e a descoberta de novos conhecimentos e lacunas por pessoas e programas de computador; fornecer um roteiro semântico para um indivíduo e promover o entendimento compartilhado, melhorar a comunicação em geral e a colaboração; apoiar a aprendizagem e a assimilação de informação; suportar a exibição significativa e bem estruturada de informações; fornecer a base conceitual para sistemas baseados em conhecimento; suportar interoperabilidade sintática e semântica e a preservação de significado ao longo do tempo; fornecer informações sobre termos, conceitos e outras entidades a leitores, escritores e tradutores para ajuda-los a compreender o texto, a conceituar um tópico e a encontrar o termo adequado.

Contudo, os métodos e as tecnologias de visualização de dados e informações e de análise de redes podem ser associados a estudos bibliométricos e cientométricos e se revelam muito eficazes em processos de recuperação da informação e de descoberta de conhecimento em *corpora* textuais.

Devido ao enorme volume de dados e informações publicado e disponível atualmente, ferramentas computacionais para o mapeamento, a organização, a representação, a análise e o reuso desses conteúdos documentais têm sido desenvolvidas e embutidas em metodologias que abrangem busca e recuperação de informações, análises quantitativas e qualitativas, categorizações e classificações de assuntos, de tópicos ou de temáticas. Assim, como as redes sociais que os produzem, dados, informações e conhecimento podem, igualmente, ser organizados em redes ou sistemas de conceitos. Esse processo envolve as premissas que conteúdos informacionais podem ser representados pelos conceitos ali incluídos, que, por sua vez, podem ser representados pelos termos que os denotam, conforme a Teoria do Conceito (Dahlberg, 1978). Finalmente, os termos e suas inter-relações num conteúdo textual, primariamente frequência e coocorrência, podem ser mapeados no formato de redes, constituindo um sistema de conceitos, inclusive com atribuições semânticas para qualificar tais relações.

O campo de visualização de informação avançou muito nos últimos anos, trazendo novas técnicas para a análise gráfica da literatura, de patentes, de genomas e outros tipos

de conteúdos informacionais (Chen & Song, 2017). No entanto, é preciso lembrar que, enquanto a visualização pode ser crítica para o entendimento e a cognição humanos, ela é simplesmente uma janela para as análises rigorosas, muitas vezes multidimensionais, que formaram a base da informática por muitos anos. Nesse sentido, o mapeamento não se refere apenas à tela de visualização, mas também às técnicas subjacentes de mineração, análise de dados e reconhecimento de padrões ou propriedades específicas dos elementos que compõem o sistema conceitual envolvido. Mapear domínios de conhecimento, então, toma como sua entrada assuntos aparentemente diversos de análises, incluindo e integrando: análises de redes (por exemplo, web, redes sociais, redes livres de escala e caminhos metabólicos); linguísticas; conceituais (domínio, conteúdo, discurso, assunto); de extração de tópicos; de citações, coautorias, coocorrência de palavras-chave; indicadores (metadados) de ciência e tecnologia; técnicas de visualização.

O principal objetivo do mapeamento de domínios de conhecimento é fornecer subsídios de análise para o especialista quanto para o não-especialista. Para o não-especialista, o mapeamento fornece um ponto de entrada em um domínio, por possibilitar a obtenção de conhecimento nos níveis macro e micro. Para o especialista, o mapeamento fornece uma validação de percepções e um meio de investigar rapidamente tendências e novas informações, incoerências e divergências, tanto quanto convergências. No entanto, até mesmo o especialista pode se surpreender com os desenvolvimentos na periferia de sua percepção. O mapeamento e a exploração interativa fornecem contextos para tais surpresas.

Vários métodos de mapeamento, organização e representação do conhecimento partem da lógica que não existe gestão do conhecimento, mas, na prática, o que se conduz é uma variedade de atividades de trabalho que viabilizem o compartilhamento de informações (Wilson, 2002). Dessa forma, enquanto representante convencional do conhecimento explicitado, o documento é a unidade de análise mais comum, porque pode ser usado para mapear um campo científico ou técnico específico e o seu desenvolvimento. Cocitação e coocorrência de termos são os dois tipos mais comuns de análise, e muitas vezes levam a diferentes agrupamentos. Nos níveis mais refinados, as técnicas de cocitação agrupam documentos por paradigma científico, ou pela mesma questão de pesquisa e hipóteses, ao passo que as comunidades de coocorrência de termos são de natureza mais tópica.

O mapeamento em redes de termos de indexação revela a estrutura cognitiva de um campo (Callon & Law, 1983). Há algum debate sobre se a análise conjunta deve ser usada para estudos de dinâmica da ciência (Leydesdorff, 1997). As abordagens mais confiáveis visam combinar técnicas de palavras coocorrentes com análises de citações (Noyons, Moed & Luwel, 1999). Técnicas mais avançadas, usando algoritmos sofisticados para agrupar e relacionar tópicos, são uma grande promessa para estudos dinâmicos (Griffiths & Steyvers, 2004; Erosheva, Fienberg & Lafferty, 2004).

### **3. Metodologia para o mapeamento conceitual**

Este item apresenta a metodologia proposta para o mapeamento conceitual de domínios de conhecimento (Moresi, Pierozzi Júnior, Oliveira, & Brandão, 2019) e no item 4 é apresentado um exemplo utilizando o tema inteligência artificial na educação.

Questões como gestão da informação e do conhecimento, descoberta de conhecimento, inteligência e planejamento estratégicos e estudos de prospecção antecipativa passaram a ser relevantes no desenvolvimento institucional das organizações, independente dos setores aos quais elas se vinculam (governamentais, acadêmicos, terceiro setor, entre outros). Essa onda de crescentes e diversos interesses, objetivos e estratégias institucionais se converte no insumo principal para a sistematização proposta no modelo descrito na Figura 1, visando o mapeamento, a organização e a representação de domínios de conhecimento específicos.

O processo se constitui de quatro etapas: mapeamento do domínio do conhecimento; codificação do conhecimento; aplicação de linguística de *corpus* e de processamento de linguagem natural; e representação do conhecimento.

O mapeamento do domínio do conhecimento se refere à definição do espaço do conhecimento humano a ser modelado no processo. No entanto, isso não significa um corte e um isolamento desse domínio de conhecimento: tal delimitação é um exercício conceitual para identificação e reconhecimento da temática a ser explorada e posterior análise de contextos, propósitos, estratégias, tendências, oportunidades, desafios, etc., vinculados a essa temática. Essa etapa pode ser executada a partir de escolha pessoal das temáticas ou envolver equipes de trabalho, por meio de dinâmicas sociais presenciais ou virtuais. Os dados podem ser coletados por meio de técnicas como *brainstormings*, entrevistas, questionários, leituras, entre outras.

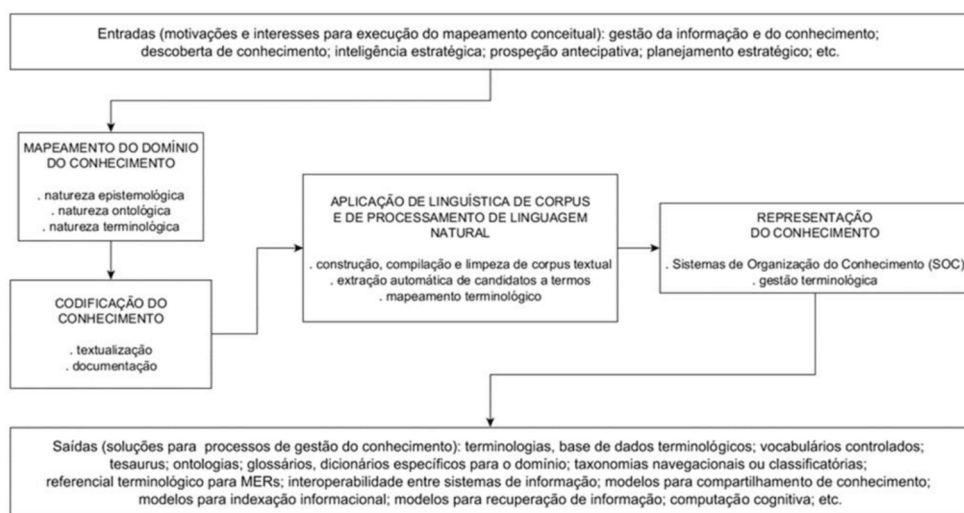


Figura 1 – Metodologia de mapeamento conceitual

A codificação do conhecimento se refere à conversão do conhecimento tácito um formato explícito, especificamente, em “textualização”, ou seja, transposição de áudios, vídeos ou qualquer outro formato em que o conhecimento tenha sido registrado, para um formato textual, reunindo o maior conteúdo possível referente às temáticas que ainda



não se apresente nessa forma como, por exemplo, as discussões socializadas em sessões de *brainstormings* ou painéis. Nessa etapa, utilizam-se técnicas diversas de registro, documentação e transcrição e são reunidos os documentos que comporão o *corpus* a ser processado na etapa que se segue.

Na sequência, as aplicações de linguística de *corpus* e de processamento de linguagem natural se realizam em três fases: construção, compilação e limpeza dos documentos selecionados para a composição do *corpus* textual para análise. Essa fase também inclui:

- a transformação de arquivos PDF ou em outros formatos, para o padrão TXT, que é o mais adequado ao processamento computacional utilizado nas fases seguintes;
- a extração semiautomática de candidatos a termos com o emprego de *softwares* contendo algoritmos para essa extração com base em métricas de frequência e coocorrência de palavras ou expressões (n-gramas) no *corpus*, utilizando o *software* VOSviewer;
- o mapeamento terminológico em que se recomenda um alinhamento terminológico dos candidatos a termos obtidos na fase anterior com recursos léxicos disponíveis e referentes às temáticas definidas para a análise.

A representação do conhecimento se propõe a produzir os artefatos explicitados textualmente. Recomenda-se novamente uma validação pelos especialistas de domínio, pois a intenção é gerar os produtos finais do processo de modelagem e análise. Duas fases podem ser executadas tanto em sequência como de forma independente uma da outra:

- desenvolvimento e aplicação de sistemas de organização do conhecimento que se refere à elaboração de um formato de representação diferente do textual linear. Dessa forma, desde formatos unidimensionais, como listas de termos, até os mais elaborados em estrutura e funcionalidade, como as ontologias, podem ser elaborados. Nessa fase, ainda, desenvolvem-se aplicações dos SOC para diversos processos de Gestão da Informação e do Conhecimento como, por exemplo, taxonomias navegacionais multifacetadas, que podem ser empregadas na construção de expressões booleanas de busca em bases de dados ou, ainda, estruturas conceituais como as redes semânticas que podem ser representadas no *software* yEd (yWorks, 2019);
- execução de gestão terminológica com o emprego de ferramentas e técnicas de gestão terminológica como, por exemplo, identificação de contextos visando definir os termos associados à temática delimitada no início do processo. O *software* livre e-Termos (Oliveira, 2009) é recomendado pois foi construído especificamente para esse fim, incluindo a operacionalização de praticamente todas as 4 etapas deste processo. Além disso, o e-Termos também atende as funcionalidades de um repositório de dados terminológicos.

Adicionalmente, alguns pontos específicos de operacionalidade do processo descrito acima, merecem um olhar mais atencioso: na escolha das tecnologias de suporte ao processo, baseados no exercício empírico de realização de mapeamento, organização e representação do conhecimento, algumas ferramentas metodológicas e alguns *softwares* revelaram-se bastante adequados.

O *software* VOSviewer tem se revelado muito satisfatório, oferecendo excelentes funcionalidades de edição e visualização de sistemas de conceitos. O bom desempenho de extração automática de termos deve ser atestada pelos especialistas de domínio que reconhecem nos termos propostos pelo *software* as formas lexicais apropriadas para os conceitos, referendando a Teoria do Conceito (Dahlberg, 1978), que embasa a concepção do processo aqui apresentado. Além disso, o *software* que se propõe a elaborar graficamente um panorama do domínio analisado, fornece métricas associadas à terminologia que viabiliza análises e abordagens relacionadas à análise de redes (centralidades, comunidades) e às geometrias do conhecimento, mais especificamente aos conceitos recentes de “espaço conceitual” (Gärdenfors, 2014) e “esfera semântica” (Lévy, 2011).

O VOSviewer, além das funcionalidades de organização e visualização de dados e informações, exporta os dados numéricos em formatos interoperáveis o que viabiliza, a partir dos arquivos de mapa de termos e de arestas para reconstruir as triplas “conceito-relação-conceito” do sistema de conceitos gerado.

A rede pode ser importada para o *software* yEd, que possibilita a representação visual original obtida no VOSviewer e a edição no formato de grafos com inclusão/exclusão de termos/conceitos (nós) e de relacionamentos (arestas) e, igualmente pode ser visualizada por meio de variados *layouts* que proporcionam novos *insights* cognitivos. Outra funcionalidade muito útil do yEd é a possibilidade de elaborar recortes do sistema de conceitos, mais uma vez motivando outros *insights* cognitivos e empoderando o processo aqui apresentado na construção de ressignificação dos dados e informações ou, em outras palavras, na construção de inteligência.

#### 4. Exemplo prático

Atualmente, a educação emprega vários métodos para ensino e aprendizagem, que utilizam tecnologias da informação e comunicação, visando adaptar e melhorar a experiência do estudante. Os softwares educacionais, por exemplo, permite a criação de inúmeras ferramentas com o objetivo de melhorar os processo de ensino e aprendizagem. Uma alternativa para a criação desse tipo de ferramentas é a inteligência artificial, que é um campo de estudos multidisciplinar, originado da computação, da engenharia, da psicologia, da matemática e da cibernética. O seu principal objetivo é construir sistemas que apresentem comportamento inteligente e desempenhem tarefas com um grau de competência equivalente ou superior ao de um especialista humano (Nikolopoulos, 1997).

Quando aplicada à Educação, os métodos de IA podem ser usados para obter e representar conhecimento em processos de ensino e aprendizagem, oferecendo as ferramentas necessárias aos professores para reunir evidências de maneira sistemática, detalhada e incremental em apoio às suas práticas pedagógicas.

Alguns fundamentos importantes para a adoção da perspectiva da contribuição metodológica da IA para a Educação foram estabelecidos Bundy (1986), que a categorizou em três tipos: (i) IA básica, com o objetivo de explorar técnicas computacionais para simular o comportamento inteligente, (ii) IA aplicada, focada em usar as técnicas existentes de IA para criar produtos para atender necessidades do mundo real e (iii) ciência cognitiva com foco no estudo da inteligência humana ou animal por meios



computacionais. Essa distinção permanece crucial na medida em que permite uma definição mais precisa da IA na educação, trazendo à tona a dependência entre as tecnologias inteligentes e educacionais projetadas e a finalidade, o contexto e o design do seu uso (Mark, & Greer, 1993).

Após esta conceituação sobre IA na Educação, foi realizada uma pesquisa na base Elsevier Scopus utilizando a expressão – *education AND “artificial intelligence”*. O objetivo foi a recuperação de forma mais abrangente. Então, a pesquisa resultou em 51.768 documentos, no período de 1970 a 2020. A Figura 2 apresenta a evolução dos quantitativos de documentos publicados no período. Observa-se que o pico de publicações ocorreu em 2017.

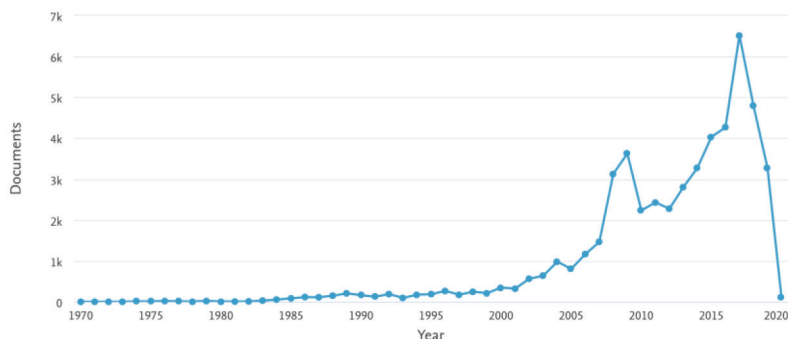


Figura 2 – Evolução das publicações sobre o tema inteligência artificial na educação

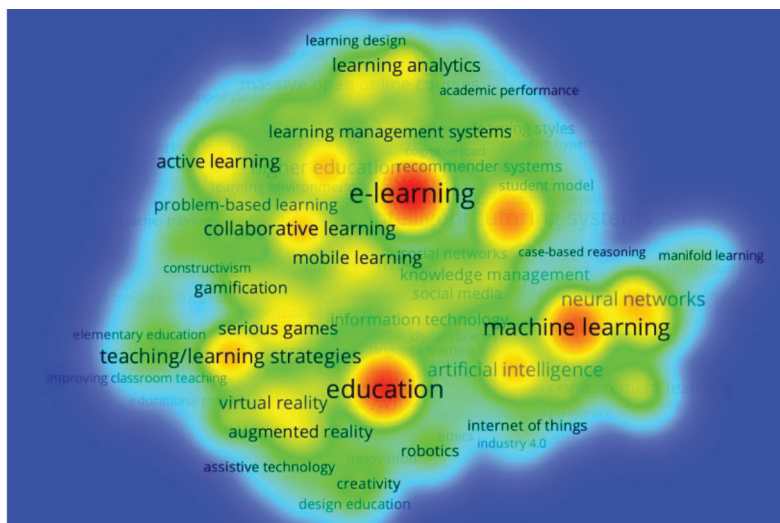


Figura 3 – Evolução das publicações sobre o tema inteligência artificial na educação.

Os metadados da pesquisa bibliográfica foram lidos pelo VOSviewer (Waltman, van Eck, & Noyons, 2010) e selecionada a opção de coocorrência de palavras-chave do autor. O software recuperou 73.533 palavras-chave. Foi selecionada a opção de ocorrência de pelo menos 50 vezes e sem emprego do arquivo de tesouros para o controle do vocabulário. A rede resultou em 359 palavras-chave, 4 *clusters* e 14.028 arestas.

Para melhorar a visualização da coocorrência de palavras foi elaborado um arquivo de Tesouros, em que os termos genéricos (metodologia, tecnologia, software, etc) foram eliminados e os termos com mesmo significados, mas com ortografias diferentes, foram atribuídos a um único termo. Por exemplo, termos no singular e no plural, foram direcionados para apenas um deles. A rede de coocorrência de palavras-chave foi gerada outra vez, mas com o arquivo de Tesouros e pelo menos 50 ocorrências, resultando em 152 termos, 5 *clusters* e 3.557 arestas. A Figura 3 apresenta uma visualização da coocorrência de termos das referências recuperadas na pesquisa bibliográfica.

Nota-se que os seguintes termos se destacam: *education, e-learning, machine learning, teaching/learning strategies, intelligent tutoring systems, artificial intelligence, active learning, collaborative learning*, entre outros. Contudo, alguns termos com baixa frequência devem ser observados por oferecerem oportunidades de pesquisa, tais como: *learning design, manifold learning, problem-based learning, robotics, augmented reality, improving classroom teching, etc.*

Para aprofundar a análise dos dados, foi usado o Gephi, um software de código aberto para exploração e manipulação de redes. A rede gerada pelo VOSviewer foi salva no formato GML para ser importado pelo Gephi, cujos módulos podem importar, visualizar, filtrar, manipular e exportar diversos tipos de redes (Bastian, Heymann, & Jacomy, 2009). Foram calculadas as métricas de grau médio, classe de modularidade e centralidade de autovetor.

A Figura 4 mostra o grafo resultante. Observa-se que alguns tópicos se destacam: *education*, *higher education*, *e-learning*, *collaborative learning*, *artificial intelligence*. As arestas com maior espessura revelam o forte relacionamento entre estes tópicos.

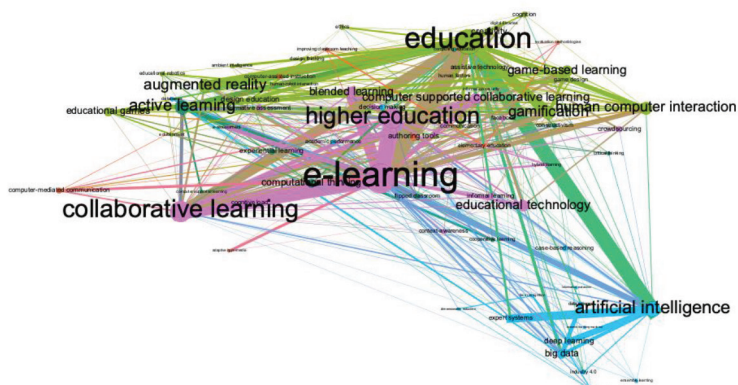


Figura 4 – Rede de coocorrência de palavras-chave representada no Gephi

Contudo, o Gephi possui uma funcionalidade de Laboratório de Dados, onde é possível extrair informações sobre as métricas de redes. A Tabela 1 apresenta as 15 palavras-chave com maiores centralidades de autovetor.. Observa-se que o termo com maior centralidade de autovetor é o *e-learning*.

<b>Termo</b>	<b>Cluster</b>	<b>Grau</b>	<b>Centralidade de Autovetor</b>
<i>e-learning</i>	1. Educational technology	148	1,0000
<i>education</i>	2. Education	130	0,9361
<i>teaching/learning strategies</i>	2. Education	128	0,9266
<i>collaborative learning</i>	1. Educational technology	125	0,9141
<i>intelligent tutoring systems</i>	4. Intelligent systems	120	0,8750
<i>higher education</i>	1. Educational technology	114	0,8588
<i>mobile learning</i>	1. Educational technology	97	0,7707
<i>machine learning</i>	3. Artificial intelligence	107	0,7646
<i>artificial intelligence</i>	3. Artificial intelligence	99	0,7636
<i>active learning</i>	5. Pedagogical issues	99	0,7624
<i>learning management systems</i>	1. Educational technology	91	0,7221
<i>human computer interaction</i>	2. Education	91	0,7220
<i>gamification</i>	2. Education	89	0,7156
<i>serious games</i>	2. Education	91	0,7125
<i>massive open online courses</i>	1. Educational technology	89	0,7089

Tabela 1 – Termos mais relevantes por centralidade de autovetor

A identificação dos *clusters* foi realizada a partir da análise dos termos em cada comunidade. A Tabela 2 apresenta os termos de cada cluster.

<b>Cluster</b>	<b>Cluster</b>
1. Educational technology	blended learning; cognitive load; collaborative learning; computer supported collaborative learning; context-awareness; crowdsourcing; e-assessment; e-learning; educational technology; facebook; formative assessment; higher education; hybrid learning; informal learning; inquiry-based learning; instructional design; knowledge sharing; learning analytics; learning design; learning environments; learning management systems; learning objects; learning outcomes; learning styles; lifelong learning; massive open online courses; mobile learning; mobile technology; open educational resources; peer assessment; personalized learning; recommender systems; social learning; social media; social networking; social networks; student performance; technology-enhanced learning; twitter; ubiquitous learning; virtual laboratory; virtual learning environments; virtual worlds; web 2.0; web-based learning.
2. Education	ambient intelligence; assistive technology; augmented reality; cognition; communication; creativity; design education; design thinking; digital libraries; education; educational games; educational robotics; edutainment; ethics; game design; game-based learning; gamification; human computer interaction; human factors; human-robot interaction; information security; information technology; innovation; interaction design; language learning; learning Technologies; mixed reality; mobile application; mobile devices; multimedia; participatory design; robotics; serious games; special education; storytelling; teacher education; teaching/learning strategies; ubiquitous computing; user experience; user interface; user-centered design; virtual reality.

Cluster	Cluster
3. Artificial intelligence	academic performance; artificial intelligence; big data; case-based reasoning; data Science; databases; decision making; deep learning; dimensionality reduction; ensemble learning; expert systems; extreme learning machine; face recognition; incremental learning ;industry 4.0; information extraction; intelligent systems; internet of things; knowledge acquisition; knowledge Discovery; knowledge engineering; knowledge management; machine learning; manifold learning; multi-agent systems; neural networks; q-learning; reinforcement learning; speech recognition; transfer learning.
4. Intelligent systems	adaptive hypermedia; authoring tools; computer-assisted instruction; computer-mediated communication; elementary education; evaluation methodologies; improving classroom teaching; intelligent tutoring systems; interactive learning environments; item response theory; media in education; metacognition; pedagogical agents; primary education; secondary education; self-regulated learning; student model; student modeling; user modeling; web-based education.
5. Pedagogical issues	active learning; computational thinking; computer supported learning; computing education; constructivism; cooperative learning; critical thinking; experiential learning; flipped classroom; pedagogical issues; problem-based learning; project-based learning; student engagement; teacher training.

Tabela 2 – Palavras-chave de cada *cluster*

Em seguida, as listas de termos e de arestas foram exportadas em formato CSV e recuperadas em uma planilha Excel, para serem analisadas no yEd<sup>2</sup>, que é um software livre para edição de grafos. Devido à limitação de representação do yEd, foram selecionadas as 30 palavras-chave mais relevantes sobre o tema e obtida a visualização dos termos *artificial intelligence* e *education*. O espaço conceitual é apresentado na Figura 5. É importante ressaltar que a inteligência artificial na educação não pode ter foco apenas na tecnologia. Deve-se entender como a tecnologia poderá sustentar novas estratégias de ensino e aprendizagem.

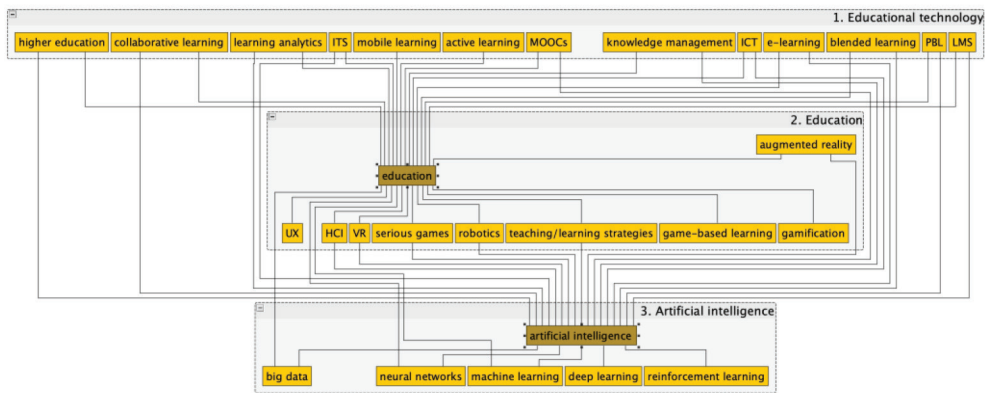


Figura 5 – Espaço conceitual do domínio temático de inteligência artificial na educação

<sup>2</sup> Disponível em: <https://www.yworks.com/products/yed>

## 5. Discussão

A exploração de um domínio do conhecimento, utilizando um espaço conceitual (Gärdenfors, 2014), possibilita a visualização de relações entre palavras-chave representadas pelos conceitos ali incluídos (Dahlberg, 1978). Para a definição da expressão de busca, é importante que o domínio temático a ser analisado seja conceituado com foco na abrangência da recuperação de referências bibliográficas.

No exemplo apresentado, foi recuperado uma grande quantidade de referências, o que tornou inviável a interpretação dos resultados sem o auxílio de ferramentas computacionais. Outro ponto a destacar é que os metadados das referências bibliográficas podem ser organizados em redes ou sistemas de conceitos. O mapeamento de coocorrência de palavras-chave, por meio de análise de redes e da utilização de vocabulário controlado, constitui um sistema de conceitos que é muito útil na análise de um domínio temático (Moresi et al, 2019).

Para o mapeamento do tema inteligência artificial e educação, foram utilizados os recursos do VOSviewer, para a leitura dos metadados dos documentos, do Gephi, para o cálculo das métricas de redes, e do yEd, para a representação do espaço conceitual do domínio temático. Os resultados mostraram que foram identificados 5 subdomínios que constituem o tema analisado: tecnologia educacional, educação, inteligência artificial, sistemas inteligentes e assuntos pedagógicos. Esse ponto é importante, porque a tecnologia deve ser vista como meio enquanto a educação como fim. Isso significa que as estratégias de ensino e aprendizagem deverão evoluir com o apoio da inteligência artificial.

## 6. Conclusões

O objetivo deste artigo foi apresentar e descrever uma metodologia de mapeamento de domínios de conhecimento utilizando ferramentas computacionais, que facilitam a elaboração de sistemas de conceitos, expressões de busca e significação da informação recuperada. O exemplo apresentado mostra a importância da abrangência de uma pesquisa bibliográfica para a recuperação de referências.

Assim, as melhorias e o aumento das opções de visualização dos dados e informações, assim como a agregação de recursos de semântica computacional, trazem novas possibilidades de *insights* cognitivos para o reconhecimento de padrões em sistemas conceituais complexos. Dessa forma, as soluções relatadas no presente artigo aumentam a aplicabilidade e o uso efetivo de dados e informações em processos envolvendo o desenvolvimento e aplicações de inovações em Ciência & Tecnologia, apontando tendências, oportunidades e desafios em contextos mais organizados de domínios do conhecimento.

Em nível mais operacional, as soluções metodológicas e tecnológicas aqui apresentadas trazem facilidades nos processos de reunião, processamento, compartilhamento e disseminação do conhecimento, explorando o potencial de linguagens híbridas e complementares de representação. Isso permite uma melhor apreensão desses conteúdos informacionais em diversos níveis de reuso.

Como perspectivas futuras de pesquisa, sugere-se que sejam exploradas outras unidades de análise bibliométrica tais como redes de citação, de cocitação e de acoplamento bibliográfico de documentos. Estes tipos de análises podem revelar fronteiras de pesquisa, bem com a estrutura cognitiva de um domínio temático.

## Referências

- Bassecoulard, E., & Zitt, M. (1999). Indicators in a research institute: A multi-level classification of scientific journals. *Scientometrics*, 44 (3), 323–345.
- Bastian, M., Heymann, S., & Jacomy, M. (2009). Gephi: An Open Source Software for Exploring and Manipulating Networks. In: *Proceedings of the Third International ICWSM Conference*, 361–362.
- Bundy, A. (1986). *What kind of field is artificial intelligence?* In: DAI Research Paper No.305. Department of Artificial Intelligence, University of Edinburgh.
- Mark, M.A., & Greer, J. (1993). Evaluation methodologies for intelligent tutoring systems. *Journal of Artificial Intelligence in Education*, 4, 129–153.
- Cahlik, T. (2000). Comparison of the maps of science. *Scientometrics*, 49(3), 373–387.
- Callon, M., & Law, J. (1983). From translations to problematic networks: An introduction to co-word analysis. *Social Science Information*, 22, 191–235.
- Chen, C., & Song, M. (2017). *Representing Scientific Knowledge: The Role of Uncertainty*. Springer International Publishing.
- Clauser, J., & Weir, S. M. (1976). *Intelligence research methodology: An introduction to techniques and procedures for conducting research in defense intelligence*. Washington: Defense Intelligence School.
- Dahlberg, I. (1978). Teoria do conceito. *Ciência da informação*, 7(2), 101–107.
- Erosheva, E., Fienberg, S., & Lafferty, J. (2004). Mixed-membership models of scientific publications. In *Proceedings of National Academy of Sciences of the United States of America*, 101, 5220–5227.
- Gärdenfors, P. (2014). *The Geometry of Meaning: Semantics Based on Conceptual Spaces*. Cambridge, MA: MIT Press.
- Garfield, E. (1994). Scientography: Mapping the tracks of science. *Current Contents: Social & Behavioural Sciences*, 7(45), 5–10.
- Griffiths, T. L., & Steyvers, M. (2004). Finding scientific topics. In *Proceedings of National Academy of Sciences of the United States of America*, 101, 5228–5235.
- Hjørland, B. (1997). *Information seeking and subject representation: an activity-theoretical approach to information science*. Westport: Greenwood Press.
- Levine, D., Strother-Garcia, K., Hirsh-Pasek, K., & Golinkoff, R. M. (2017). Names for things... and actions and events: following in the footsteps of Roger Brown. In Fernández, E. M., & Cairns, H. S. (Ed.) (2017). *The Handbook of Psycholinguistics*. Hoboken, NJ: John Wiley & Sons, Inc.



- Lévy, P. (2011). *The semantic sphere: computation, cognition, and information economy*. Hoboken, NJ: John Wiley & Sons, Inc.
- Leydesdorff, L. (1997). Why words and co-words cannot map the development of the sciences. *Journal of American Society of Information Science*, 48, 418–427.
- Moresi, E. A .D., Pierozzi Júnior, I., Oliveira, L. H. M., & Brandão, A. M. (2019). Organização e representação de conhecimento: incrementos metodológicos e tecnológicos para o mapeamento conceitual. Atas do 8o Congresso Ibero-Americano em Investigação Qualitativa 2019, Lisboa, Portugal, 3, 269–278.
- Nikolopoulos, C. (1997). *Expert systems: introduction to first and second generation and hybrid knowledge based systems*. New York: Marcel Dekker, Inc.
- Noyons, E. C. M., Moed, H. F., & Luwel, M. (1999). Combining mapping and citation analysis for evaluative bibliometric purposes: a bibliometric study. *Journal of American Society of Information Science*, 50, 115–131.
- Oliveira, L. H. M. (2009). *E-TERMOS: Um ambiente colaborativo web de gestão terminológica*. (Tese de Doutorado em Ciência da Computação), Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Paulo.
- Salvador, M. R., & Lopez-Martinez, R. E. (2000). Cognitive structure of research: Scientometric mapping in sintered materials. *Research Evaluation*, 9, 189–200.
- Shiffrin, R. M., & Börner, K. (2004). Mapping knowledge domains. In *Proceedings of the National Academy of Sciences of the United States of America*, 101 (Suppl 1), 5183–5185.
- Soergel, D. (2008). Digital Libraries and Knowledge Organization Systems. In Kruk, S. (Ed.) (2008). *Semantic Digital Libraries*. Berlin: Springer.
- Torres, T. Z., Pierozzi Jr, I., Pereira, N. R., & De Castro, A. (2011). Knowledge management and communication in Brazilian agricultural research: An integrated procedural approach. *International Journal of Information Management*, 31(2), 121–127.
- Waltman, I., van Eck, N. J., & Noyons, E. C. M. (2010). A unified approach to mapping and clustering of bibliometric networks. *Journal of Informetrics*, 4 (4), 629–635.
- Wilson, T. D. (2002). The nonsense of knowledge management. *Information Research*, 8(1), 1–8.
- Yworks (nd). yEd Graph Editor: High-quality diagrams made easy. Retrieved from <https://www.yworks.com/products/yed>.
- Zhao, D., & Strotmann, A. (2008). Evolution of research activities and intellectual influences in information science 1996–2005: Introducing author bibliographic-coupling analysis. *Journal of the American Society for Information Science and Technology*, 59 (13), 2070–2086.