



UNIVERSIDADE ESTADUAL DO MARANHÃO
CENTRO DE CIÊNCIAS TECNOLÓGICAS
PROGRAMA DE POS-GRADUAÇÃO EM ENGENHARIA DA COMPUTAÇÃO E
SISTEMAS

LEILANNE FRANCISCA DIAS BORGES

**DESENVOLVIMENTO DE FERRAMENTA DE SISTEMATIZAÇÃO PARA
ACOMPANHAMENTO DE AUTOESTUDO POR MEIO DE RECOMENDAÇÃO DE
INFORMAÇÃO BASEADA EM CONTEXTO**

São Luís

2020

LEILANNE FRANCISCA DIAS BORGES

**DESENVOLVIMENTO DE FERRAMENTA DE SISTEMATIZAÇÃO PARA
ACOMPANHAMENTO DE AUTOESTUDO POR MEIO DE RECOMENDAÇÃO DE
INFORMAÇÃO BASEADA EM CONTEXTO**

Trabalho apresentado ao Programa de Pós -
Graduação em Engenharia da Computação da
Universidade Estadual do Maranhão, como
parte dos requisitos para obtenção do título de
Mestre em Engenharia da Computação e
Sistemas.

Orientador: Prof. Dr. Luís Carlos Costa
Fonseca

São Luís

2020

Borges, Leilanne Francisca Dias.

Desenvolvimento de ferramenta de sistematização para acompanhamento de autoestudo por meio de recomendação de informação baseada em contexto / Leilanne Francisca Dias Borges. – São Luís, 2020.

63 f

Dissertação (Mestrado) – Curso de Engenharia da Computação e Sistemas, Universidade Estadual do Maranhão, 2020.

Orientador: Prof. Dr. Luís Carlos Costa Fonseca.

1.Aprendizagem continuada. 2.Projetos de aprendizagem. 3.Recuperação de informação. 4.Recomendação de informação. 5.Autoestudo. I.Título.

CDU: 004.4:37

LEILANNE FRANCISCA DIAS BORGES

**DESENVOLVIMENTO DE FERRAMENTA DE SISTEMATIZAÇÃO PARA
ACOMPANHAMENTO DE AUTOESTUDO POR MEIO DE RECOMENDAÇÃO DE
INFORMAÇÃO BASEADA EM CONTEXTO**

Trabalho apresentado ao Programa de Pós -
Graduação em Engenharia da Computação da
Universidade Estadual do Maranhão, como
parte dos requisitos para obtenção do título de
Mestre em Engenharia da Computação e
Sistemas.

Orientador: Prof. Dr. Luís Carlos Costa
Fonseca

Trabalho aprovado. São Luís, 30 de setembro de 2020.

Dr. Luís Carlos Costa Fonseca
Orientador

Dr. Reinaldo de Jesus da Silva
Primeiro membro da banca

Dr. Othon de Carvalho Bastos Silva, UFMA
Segundo membro da banca

AGRADECIMENTOS

Primeiramente agradeço a Deus, pois sem ele nada disso seria possível.

Aos meus pais por sempre acreditarem em mim e terem me apoiado durante toda a minha jornada até este momento.

A minha irmã que sempre me ajudou e esteve ao meu lado em toda a minha vida.

Ao meu orientador professor Dr. Luís Carlos pela credibilidade e confiança a mim fornecidas na execução desse trabalho. Além da sua paciência e motivação nos momentos de dificuldade.

A todo corpo docente e discente da turma do Mestrado de Engenharia de Computação e Sistemas da UEMA pela assistência e conhecimento oferecido.

Por fim, a todos que, direta e indiretamente contribuíram para a realização deste trabalho.

“Comece fazendo o que é necessário, depois o que é possível, e de repente você estará fazendo o impossível.”

São Francisco de Assis

RESUMO

Neste trabalho é desenvolvido um projeto de uma ferramenta Web que sirva como mediadora no apoio à aprendizagem continuada por meio da recomendação em contexto. A ideia principal desta ferramenta é que ela seja utilizada por usuários autônomos que estejam em busca de construir o seu conhecimento por meio do autoestudo. Assim, a ferramenta proposta segue a metodologia pedagógica de projetos de aprendizagem, onde o aluno é o responsável pela construção de seu próprio conhecimento. Nesta metodologia, o aluno deve realizar inicialmente o levantamento de certezas e dúvidas do tema que ele busca aprender, de forma que a partir da investigação, suas dúvidas possam se tornar certezas ou ainda as certezas possam se tornar dúvidas. Deste modo, o propósito dessa ferramenta será realizar a filtragem dos termos das anotações do usuário utilizando-se de técnicas de recuperação de informação e submetê-las a motores de busca a fim de realizar a recomendação dos documentos que se adequem a sua necessidade auxiliando-o nesse processo de construção do conhecimento.

Palavras-chave: Aprendizagem continuada. Projetos de aprendizagem. Recuperação de informação. Recomendação de informação. Autoestudo.

ABSTRACT

In this work, a project is developed for a Web tool that serves as a mediator in supporting lifelong learning through information recommendation in context. The main objective of this tool is that it could be used by independent users who are looking for building their knowledges through of self-study. Thus, the proposed tool follows the pedagogical methodology of learning projects, where the student is responsible for building their own knowledge. In this methodology, the student should initially establish their certainties and doubts about the subject he seeks to learn, so that from the investigation, their doubts could become certainties or even the certainties could become doubts. Therefore, the purpose of this tool will be to filter the terms of user annotations using information retrieval techniques and submit them to search engines in order to make the recommendation of the documents that suit their needs by helping them in the process of their knowledge construction.

Keywords: Lifelong learning, Learning projects, Information retrieval, Recommender System, Self-study

LISTA DE ACRÔNIMOS

BD	<i>Banco de Dados</i>
BSON	<i>Binary JSON</i>
CAP/UFGRS	<i>Projeto Amora/UFGRS</i>
FMI	<i>Fundo Monetário Internacional</i>
HTML5	<i>Hypertext Markup Language, versão 5</i>
IDF	<i>Inverse Document Frequency</i>
IMF	<i>International Monetary Fund</i>
JSON	<i>JavaScript Object Notation</i>
LEC/UFGRS	<i>Laboratório de Estudos Cognitivos/UFGRS</i>
LLCQ	<i>Lifelong Learning Council Queensland</i>
MB	<i>Megabyte</i>
MEC	<i>Ministério da Educação</i>
NLTK	<i>Natural Language Toolkit</i>
NOSQL	<i>Not Only SQL</i>
OA	<i>Objeto de Aprendizagem</i>
OECD	<i>Organization for Economic Cooperation and Development</i>
PLE	<i>Personal Learning Environments</i>
PROINFO	<i>Programa de inserção das tecnologias da informação e da comunicação</i>
RI	<i>Recuperação de Informação</i>
SGA	<i>Sistema de Gestão de Aprendizagem</i>
SGBD	<i>Sistemas Gerenciadores de Banco de Dados</i>
SR	<i>Sistema de Recomendação</i>
TF	<i>Term Frequency</i>
UFRGS	<i>Universidade Federal do Rio Grande do Sul</i>
UNESCO	<i>Organização das Nações Unidas para a Educação, a Ciência e a Cultura</i>
WDR	<i>World Development Report - Relatório sobre o Desenvolvimento Mundial</i>

LISTA DE FIGURAS

Figura 1 - Crescimento semestral da quantidade de domínios	14
Figura 2 - Representação do processo de recuperação de informação	28
Figura 3 - Interação do usuário com o sistema de recuperação através de diferentes tarefas .	29
Figura 4 - Visões lógicas intermediárias utilizáveis para sistema de RI.....	31
Figura 5 - Taxonomia de modelos de RI.....	32
Figura 6 - Documentos Indexados com os pesos de cada documento.....	34
Figura 7- Ilustração do Modelo Vetorial.....	35
Figura 8 - Filas de prioridade dos modelos de recomendação	45
Figura 9 – Filas de prioridade dos modelos de recomendação do protótipo.....	46
Figura 10 - Estrutura do Banco de Dados MongoDB.....	48
Figura 11 – Coleção	49
Figura 12 – Quadro de Modelo de Dados	51
Figura 13 - Modelo de dados do sistema proposto	53

LISTA DE TABELAS

Tabela 1- Abordagens de aprendizagem tradicional e construtivista	21
Tabela 2 – Exemplificação de documentos JSON.....	49
Tabela 3 – Verificação da Precisão de Recomendações	56
Tabela 4 – Verificação da Precisão de Itens Recuperados.....	56

LISTA DE EQUAÇÕES

Equação 1 - Cálculo de similaridade no modelo booleano	35
Equação 2 - Cálculo TF.....	43
Equação 3 - Cálculo IDF.....	43
Equação 4 - Cálculo da Precisão	55

SUMÁRIO

1	INTRODUÇÃO.....	12
1.1	Motivação	13
1.2	Objetivos	15
1.3	Trabalhos Correlatos.....	15
1.4	Organização do Texto.....	16
2	ASPECTOS PEDAGÓGICOS.....	17
2.1	Lifelong Learning	17
2.2	Aprendizado Autônomo.....	18
2.2.1	Heutagogia	19
2.2.2	Construtivismo	20
2.3	Projetos de Aprendizagem	22
2.3.1	Metodologia de Projetos de Aprendizagem.....	23
3	AMBIENTES PESSOAIS DE APRENDIZAGEM.....	26
4	RECUPERAÇÃO DE INFORMAÇÃO	28
4.1	A Visão Lógica de um Documento	30
4.2	Modelos de Recuperação de Informação.....	31
4.2.1	Modelo Booleano	33
4.2.2	Modelo Vetorial	34
4.2.3	Modelo Probabilístico.....	36
5	SISTEMAS DE RECOMENDAÇÃO.....	38
5.1	Filtragem Baseada em Conteúdo.....	39
5.2	Filtragem Colaborativa	40
5.3	Modelo Híbrido	41
6	DESCRIÇÃO DO SISTEMA	42
6.1	Definição do Método de Recuperação de Informação	42
6.2	Definição do Sistema de Recomendação.....	44

6.3	Definição de Tecnologias	46
6.2.1	Banco de Dados.....	46
6.2.1.1	MongoDB.....	47
6.2.1.2	Modelo de Dados.....	50
6.2.2	Linguagem de Programação.....	52
6.4	Arquitetura do Sistema	53
6.5	Resultados dos Testes	54
7	CONSIDERAÇÕES FINAIS.....	57
7.1	Trabalhos Futuros.....	58
	REFERÊNCIAS	59
	APÊNDICE A – Dados de Usuário Usados Para Testes	64
	APÊNDICE B – Descrição Casos de Testes	68
	APÊNDICE C – Telas do Protótipo Desenvolvido	76

1 INTRODUÇÃO

A busca por conhecimento é uma atividade comum na sociedade. Hoje em dia, com a evolução das tecnologias computacionais e disseminação de conteúdos na Internet, torna-se obsoleto limitar-se apenas ao conteúdo repassado em sala de aula. De acordo com Flores (1996) a informática deve ser um complemento de conhecimento de conteúdos curriculares visando o desenvolvimento integral do indivíduo, habilitando e dando oportunidades ao mesmo de forma que ele adquira novos conhecimentos.

Visto o avanço das tecnologias computacionais e pela forma rápida e fluída em que as informações são disseminadas hoje em dia com o auxílio da internet, é tangível o aumento de conteúdos e informações que podemos encontrar nesse meio, de forma que conhecimentos profissionais, habilidades ou técnicas acabem se tornando defasadas de modo mais acelerado, em virtude do nascimento e disseminação de um novo método para as mesmas.

Essa volatilidade em técnicas encontradas hoje em dia desencadeia uma constante corrida de profissionais em busca de se manterem sempre atualizados. Consoante com Diniz (2017), uma forma de um profissional se manter atualizado dado as constantes mudanças encontradas decorrentes da revolução digital, é passar a compreender a aprendizagem como um projeto a longo prazo que começa na primeira infância e não tem prazo para acabar. Esse processo também é conhecido pelo nome de *lifelong learning*, ou aprendizado ao longo da vida.

O aprendizado para a atualização de conteúdos ocorre frequentemente de forma autônoma, pela internet, por meio de vídeos, tutoriais, artigos, entre outros recursos. Seguindo a mesma linha de pensamento, na área da psicologia, Ausubel (1987, apud. SARDO, 2003) descreve que o aprendizado autônomo envolve descobertas de modo independente para o estudante antes que elas sejam incorporadas na sua estrutura cognitiva.

Nessa perspectiva, a tecnologia pode ser utilizada como uma ferramenta de auxílio ao autoestudo, de modo a prover acesso a uma gama de conhecimentos atuais sobre determinado tópico em questão, assessorando o estudante na progressão de seu aprendizado ao longo da vida.

No entanto, de acordo com Fonseca (2009), um dos problemas associados na busca de conteúdos da internet que persiste até os dias atuais é o “problema de acesso à informação”, caracterizado pela dificuldade em estimar a quantidade de documentos na Web, visto a sua natureza dinâmica, onde novos documentos surgem e somem diariamente.

Esse problema foi diminuído a partir da evolução dos motores de busca (como Google, Bing, Yahoo, etc.), entretanto, nem sempre o usuário escolhe as palavras-chave mais adequadas, não conseguindo externar o que realmente procura. Outra dificuldade é que os motores de busca não guardam quaisquer informações sobre as necessidades dos usuários, não considerando a sua vontade. (FONSECA,2009)

Para que houvesse uma recuperação de informação satisfatória ao estudante enquanto realiza o seu autoestudo em determinada área, ele deveria ter acesso a uma ferramenta computacional que utilizasse como base em suas buscas anotações e vestígios do usuário. Assim, a ferramenta proposta busca auxiliar o aluno autônomo em sua constante procura por aprendizado, por meio da recuperação e recomendação de conteúdos de acordo com os vestígios deixados pelo estudante enquanto realiza seu autoestudo, como um caderno inteligente por exemplo, seguindo a metodologia pedagógica de projetos de aprendizagem, onde o aluno é o responsável pela construção de seu próprio conhecimento.

Nesta metodologia, o aluno deve realizar inicialmente o levantamento de certezas e dúvidas do tema que ele busca aprender, de forma que a partir da investigação, suas dúvidas possam se tornar certezas ou ainda as certezas possam se tornar dúvidas.

1.1 MOTIVAÇÃO

Vivemos em um mundo no qual a sociedade atingiu o mais alto grau de evolução da técnica, sendo mais satisfatório continuar evoluindo em função da tecnologia. Nessa perspectiva e corroborando com o pensamento de Freire (1996) “ensinar não é transferir conhecimento, mas criar as possibilidades para a produção ou a sua construção”. A Informática na Educação insere-se nesse contexto, por ser capaz de possibilitar que o processo ensino-aprendizagem se torne algo mais dinâmico e moderno.

De acordo com Flores (1996) a informática deve ser um complemento de conhecimento de conteúdos curriculares visando o desenvolvimento integral do indivíduo, habilitando e dando oportunidades ao mesmo de forma que ele adquira novos conhecimentos. Tal pensamento, ganha mais força hoje em dia, visto a rapidez com que a tecnologia tem evoluído, forçando com que profissionais busquem se manter cada vez mais atualizados. Essa busca por atualização de conteúdos ocorre frequentemente de forma autônoma, pela internet, por meio de vídeos, tutoriais, artigos, entre outros recursos.

Porém, a quantidade de conteúdos depositados na rede pode dificultar o trabalho de um usuário que busca um estudo autodidata. Conforme dados divulgados pelo Internet World Stats (2020), o número de pessoas com acesso à internet no Brasil chegou a 149 milhões em dezembro de 2019, representando cerca de 70,1% da população do país do mesmo ano, o que mostra uma tendência crescente do número total de pessoas com acesso à Internet. Tais dados demonstram um gradativo do número de pessoas no Brasil que possuem acesso e depositam dados na rede.

Outro dado estatístico que pode ser verificado é o crescimento do número de domínios (nomes de sites) cadastrados no período entre janeiro de 1994 e Janeiro de 2019:

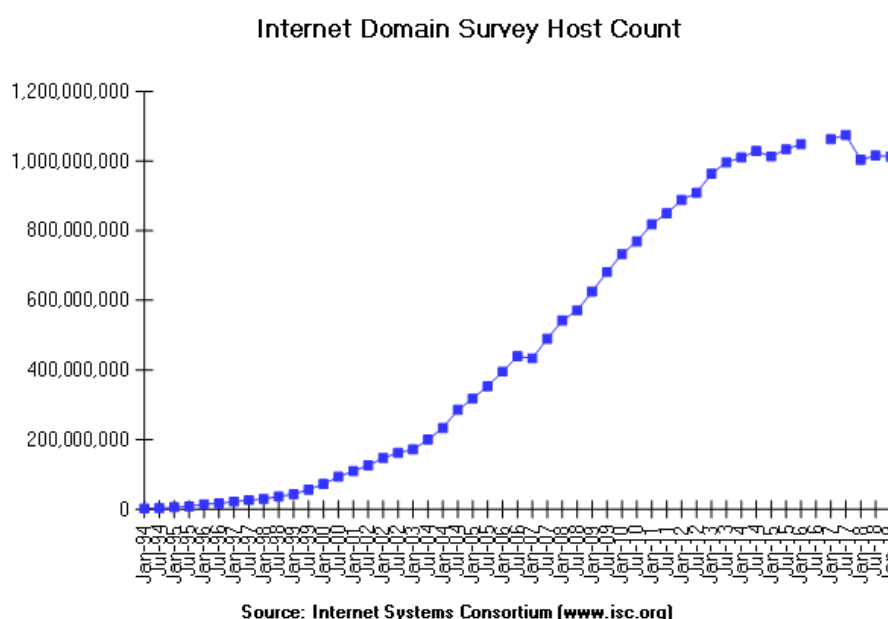


Figura 1 - Crescimento semestral da quantidade de domínios
Fonte: Internet Systems Consortium (2019)

Apesar do grande volume de dados disponíveis na internet e de sites especializados em busca, muitas vezes as pessoas levam muito tempo para localizar uma informação desejada ou nem chegam a encontrá-la.

O estudo de Cazella, Drumm e Barbosa (2010) abordam esse problema, mencionando que o usuário não é capaz de selecionar as informações mais adequadas à sua necessidade, seja por desconhecimento do tema, ou pela sobrecarga de informação provocada pelas ferramentas de busca utilizadas. Esse problema foi diminuído a partir da evolução dos motores de busca (como Google, Bing, Yahoo, etc.), entretanto, nem sempre o usuário escolhe as palavras-chave mais adequadas, não conseguindo externar o que realmente procura.

Sendo assim, este trabalho propõe a criação de um ambiente pessoal de aprendizagem que auxilie o usuário com perfil autônomo durante seu autoestudo. Para isso, o sistema proposto deverá entender as necessidades do usuário a partir das anotações de seu estudo, de tal forma que ele possa recuperar e recomendar informações de acordo com contexto em que ele está inserido.

1.2 OBJETIVOS

O objetivo principal deste projeto é desenvolver uma ferramenta que pode ser comparada a um caderno inteligente, ou seja, uma ferramenta do qual o usuário poderá depositar suas informações de estudo e ser auxiliado por meio da recomendação de informação com base das suas anotações e exposição de suas dúvidas e certezas, ao utilizar a pedagogia de Projeto de Aprendizagem.

Para tanto, estabelece-se os seguintes objetivos específicos:

- Realizar estudo comparativo com outras ferramentas similares;
- Desenvolver mecanismos para recuperar informações da internet baseadas no contexto das anotações realizadas pelo usuário;
- Investigar tecnologias que suportem os mecanismos de busca de modo que a extração de termos seja condizente com as necessidades apresentadas pelo usuário;
- Desenvolver modelo do sistema (protótipo para validação);
- Realizar experimentos com as ferramentas desenvolvidas (protótipo) por meio de casos de uso, para verificar sua eficiência.

1.3 TRABALHOS CORRELATOS

Vários outros trabalhos já realizaram estudos e propostas de ferramentas e técnicas de recomendação na área educacional, como Farias (2015), que propõe uma técnica de recomendação voltada para recomendação de links para o fomento de discussões em fóruns de um ambiente virtual de aprendizagem. Esse trabalho explana a metodologia para identificação dos principais tópicos discutidos em um fórum por meio da utilização de técnicas de mineração

de texto. Após identificação dos tópicos, os termos são submetidos a um motor de busca e retornado ao usuário os 10 (dez) primeiros links como sugestão de leitura.

Já Ferneda e Dias (2017) explana sobre a proposta de criação de um sistema de recuperação de informação denominado OntoSmart que realiza a recuperação de informação pelo modelo vetorial, se utilizando do meio de ontologias para padronização do vocabulário e representação dos documentos da busca.

Enquanto Silva (2016), propõe um sistema computacional denominado de RECOAWS que realiza a recomendação de objetos de aprendizagem baseado em postagens extraídas de dentro de fóruns de um ambiente virtual de aprendizagem, se utilizando de webservices como meio de extração das informações necessárias para recomendação, por meio dos logs de usuários.

Fonseca (2009) propõe um sistema, denominado ARPA, que realiza recuperação e recomendação de informação como auxílio para autoestudo por meio da pedagogia de Projetos de Aprendizagem se utilizando de algoritmos genéticos para recomendação de informação.

Já o presente trabalho, propõe a criação de um sistema que, além do auxílio durante o autoestudo do usuário por meio de recomendação de informações baseadas nos vestígios de informações deixadas pelo estudante durante seu autoestudo realizado por meio da pedagogia de Projetos de Aprendizagem, pode ser conectado por redes sociais, propondo uma fila de recomendação, além da implementação em bancos de dados não relacionais, dando flexibilidade necessária para constantes mudanças que possam ocorrer em no sistema.

1.4 ORGANIZAÇÃO DO TEXTO

Este trabalho está construído em sete capítulos dispostos da seguinte forma: Os conceitos referentes a abordagem pedagógica do qual este trabalho se baseia, como Lifelong Learning, Aprendizado autônomo e Projetos de Aprendizagem são apresentados no segundo capítulo. O terceiro capítulo faz uma breve abordagem sobre os Ambientes Pessoais de Aprendizagem. No quarto e quinto capítulo são expostos tópicos tecnológicos referente a metodologia utilizada no trabalhado, abordando a Recuperação de Informação (RI), destacando seus modelos e processo de ponderação de termos e os Sistemas de Recomendação (SR), descrevendo seus tipos de filtragem. O sexto capítulo delinea a descrição do sistema e é feita a demonstração da arquitetura da ferramenta. Por fim, no sétimo capítulo são apresentadas as conclusões finais deste trabalho.

2 ASPECTOS PEDAGÓGICOS

Seguindo a premissa de que estamos propondo neste trabalho uma ferramenta que auxilie no autoestudo, torna-se necessária uma abordagem pedagógica que surja como alicerce para construção da ferramenta sugerida.

Desse modo, serão explorados termos como *lifelong learning* (aprendizagem ao longo da vida), aprendizado autônomo, abordando também a *heutagogia* (aprendizado autodeterminado) e a pedagogia de projetos de aprendizagem, sendo esta última a qual o sistema propõe se guiar para ofertar um auxílio ao autoestudo ao estudante de perfil autônomo.

2.1 LIFELONG LEARNING

O conceito da *lifelong learning* (ou aprendizado ao longo da vida) como uma estratégia educacional surgiu no final do século passado devido a algumas ações de grandes organizações mundiais, como a OECD (Organization for Economic Cooperation and Development), que alterou o termo *recurrent education* (*educação recorrente*), utilizado por vários anos, pelo termo “learning” (aprendizado) e a UNESCO (Organização das Nações Unidas para a Educação, a Ciência e a Cultura) que denotou “Os anos de desenvolvimento humano sustentável (*lifelong learning*) e paz” como seu plano de médio prazo entre 1995 e 1998 (LONGWORTH; DAVIES, 2013).

Mas o que é o *lifelong learning*? De acordo com a Lifelong Learning Council Queensland (LLCQ), o *lifelong learning* pode ser amplamente definido como uma aprendizagem que é seguida ao longo da vida: uma aprendizagem flexível, diversificada e disponível em diferentes lugares e momentos. Capaz de atravessar setores, promovendo a aprendizagem além da educação tradicional e ao longo da vida adulta, ou seja, uma educação pós-obrigatória, além da educação que recebemos nas escolas e universidades.

Em um mundo que se encontra em constantes mudanças, a aplicação do *lifelong learning* pode ser vista como uma segunda chance de atualização de habilidades para que profissionais possam continuar atualizados em seu ambiente de trabalho.

A WDR (World Development Report - Relatório sobre o Desenvolvimento Mundial) de 2019 traz a *lifelong learning* como um de seus capítulos e abre a discussão com uma ampla variedade de casos de desenvolvimentos positivos no uso da tecnologia, como em

robôs, sites e inteligência artificial. O mesmo relatório ainda afirma que a tecnologia tem criado mais empregos do que desempregos (STROMQUIST, 2019).

Já um documento de trabalho do FMI (Fundo Monetário Internacional - International Monetary Fund – IMF) de 2018, menciona que vários formuladores de políticas e economistas veem a educação como a principal solução para o desemprego devido à tecnologia. O mesmo documento afirma que a educação deve ser vista como um investimento para converter trabalhadores de “não qualificados” para “qualificados”. Incontestavelmente, isso reduziria a desigualdade salarial e fortaleceria a demanda por mão de obra qualificada (BERG; BUFFIE; ZANNA, 2018).

Neste ambiente apresentado em que existe a necessidade de possuir um aprendizado continuado ao longo da vida, muitos profissionais buscam se atualizar de forma autodidata por meio da internet. No entanto, é necessário que o aluno desenvolva sua autonomia, se tornando mais ativo e independente. De acordo com Silvia et al. (2016), para desenvolver tal autonomia é preciso mudar a visão empirista de treino e prática, para uma visão construtivista de resolução de problemas, com o objetivo de favorecer a aprendizagem. A seguir, será explanado um pouco mais sobre o aprendizado autônomo.

2.2 APRENDIZADO AUTÔNOMO

Neste trabalho será mostrada uma ferramenta web que auxiliará o estudante durante seu autoestudo. No entanto, para que o estudante consiga construir seu conhecimento por meio desta ferramenta é necessário que ele desenvolva sua autonomia, se tornando mais ativo e independente.

Mas como podemos definir a autonomia? Etimologicamente o termo autonomia vem do grego *autós* (próprio, a si mesmo) *enomos* (lei, norma, regra) e significava a capacidade de cada cidade governar e definir seus preceitos e leis, e dos cidadãos decidirem o que fazer (PRETI, 2000, apud SANTOS, 2015).

A autonomia está relacionada ao próprio indivíduo, à sua capacidade de buscar por si mesmo, sem uma dependência explícita de um outro indivíduo. Pedagogicamente, a autonomia configura-se em reconhecer no outro a capacidade de participar, ter o que oferecer e poder decidir em conjunto com o potencial do sujeito em ser o responsável pela sua própria formação (SANTOS, 2015). Para Freire (2000, apud SANTOS, 2015), a autonomia do ser é um princípio que se constitui como fundamental a prática educativa.

Segundo Belloni (1999, apud SANTOS, 2015), no processo de aprendizagem autônoma o estudante não pode ser definido como objeto ou produto, mas como o sujeito ativo que realiza sua própria aprendizagem e abstrai o conhecimento adquirido aplicando-o em novas situações. Ainda de acordo com Belloni (2006, apud PINO; RESENDE, 2015), o estudante virtual do século XXI deve ter autonomia para aprender a gerir situações de grupo, compreender novas situações, com maior mobilidade e ser responsável por sua autoaprendizagem.

De tal modo, podemos compreender que para que o estudante possa ser considerado autônomo, ele deve possuir habilidades que facilitem em seu autoaprendizado. Essa autonomia pode ser remetida a heutagogia, ou aprendizagem autodeterminada.

2.2.1 HEUTAGOGIA

O termo heutagogia vem do grego *heuta* (auto) e *agogus* (guiar) e trata-se de um conceito que expande a concepção de andragogia (aprendizagem de adultos) ao reconhecer as experiências cotidianas como fonte de saber, incorporando a autodireção da aprendizagem com foco nas experiências (ALMEIDA, 2003, apud COELHO; DUTRA; MARIELI, 2016).

Esse termo foi criado em 2000 pelos professores Stewart Hase e Kenyon da Southern Cross University - Austrália. Para eles a heutagogia é um paradigma educacional que propõe um processo do qual o professor é o provedor de recursos, mas cabe ao estudante buscar seu percurso de aprendizagem, determinando o que e como aprender (PINO; RESENDE, 2015).

De acordo com Almeida (2003, apud COELHO; DUTRA; MARIELI, 2016), na heutagogia é contemplado o fato de o aluno administrar sua própria aprendizagem, com flexibilidade, delimitando as formas e modelos comportamentais que facilitam a sua busca pelo conhecimento.

Segundo Pino e Resende (et al., 2016):

A Heutagogia ou aprendizado autodeterminado, como afirmam Blaschke, Hase e Kenyon (2014), não é algo que fazemos às pessoas. Os indivíduos estão pré-dispostos a aprender. A aprendizagem autodeterminada surge a partir de uma experiência otimizada com o acesso à informação por meio de diversos dispositivos tecnológicos comunicacionais. Foca a Educação em aspectos cognitivos e nas necessidades das pessoas do século XXI.

A heutagogia está relacionada à noção filosófica do autodeterminismo, compartilhando a crença comum a respeito do comportamento da humanidade sobre informações e conhecimento favorecidos pela tecnologia e acesso à rede (PINO; RESENDE, 2015).

Desse modo, no aprendizado autônomo com a abordagem heutagógica, entende-se que existe a necessidade de uma flexibilização no processo de ensino-aprendizagem, onde o professor é colocado como recurso do estudante, mas é o estudante o responsável por delinear seu percurso durante o aprendizado (COELHO; DUTRA; MARIELI, 2016).

Sendo assim, o modelo heutagógico de ensino sugere um aumento da autonomia do estudante, de modo que o mesmo seja capaz de definir seu modelo de estudo, escolhendo *como* e *o que* aprender de acordo com os seus objetivos, sendo capaz de construir seu conhecimento a partir de sua própria experiência.

Dessa forma, a heutagogia apresenta alguns pontos que convergem para a proposta construtivista, em que as pessoas constroem seu próprio senso de mundo e realidade, de acordo com suas próprias experiências, de modo que o estudante se encontre ativamente envolvido em sua própria aprendizagem. A seguir será realizada uma explanação sobre o construtivismo.

2.2.2 CONSTRUTIVISMO

Seguindo a premissa de que estamos propondo neste trabalho uma ferramenta que auxilie no autoestudo, torna-se imprescindível abordar sobre teorias de modelos pedagógicos que surgem como embasamento para construção da ferramenta sugerida. Sendo assim, segue-se uma abordagem sobre o modelo construtivista e a seguir será abordada a metodologia de projetos de aprendizagem.

De acordo com Arias e Yera (2012), o construtivismo não surgiu como uma teoria pedagógica, mas como um conceito filosófico-psicológico sobre o desenvolvimento mental do ser humano.

Segundo Piaget (1929-1979), o fenômeno da aprendizagem depende de um processo construtivo que ocorre por meio de construções e reconstruções dos sistemas de significação e dos sistemas lógicos de cada indivíduo. Para que isso ocorra, é fundamental que esse indivíduo possa interagir com os objetos (no sentido piagetiano: natureza, mundo físico, cultura, artes, ciências, linguagens...), com outros sujeitos (sociedade, instituições...) e, agora, com tecnologias. [...] (FONSECA, 2009).

No entanto, no âmbito educacional, o construtivismo concebe a aprendizagem como um processo de construção dos conhecimentos, de sua elaboração, de diálogo, onde o epicentro desse processo é o próprio estudante e na forma em que ele constrói e reconstrói o seu próprio saber (ARIAS; YERA, 2012).

De acordo com Jonassen (1996):

[...] As concepções tradicionais de aprendizagem admitem que o conhecimento é um objeto, algo que pode ser transmitido do professor para o aluno. Esta concepção presume que o conhecimento é algo que pode ser adquirido, como suprimentos comprados num supermercado. Os construtivistas, por outro lado, acreditam que o conhecimento é uma construção humana de significados que procura fazer sentido do seu mundo. Os seres humanos são observadores e intérpretes naturais do mundo físico. A fim de realizar isto, eles explicam ideias e fenômenos novos nos termos do conhecimento existente.

Essas diferenças entre as concepções tradicionais e construtivistas podem ser mais bem visualizadas na tabela 1, onde podemos visualizar uma comparação entre ambas:

Abordagem Tradicional	Abordagem Construtivista
Enfoque no professor	Enfoque no aluno
Enfoque no conteúdo	Enfoque na construção individual de significados
A mente do aluno funciona como uma “tabula rasa”	A aprendizagem é uma construção do aluno sobre conhecimentos prévios.
O aluno é receptor passivo de conhecimento.	Ênfase no controle do aluno sobre sua aprendizagem
Memorização de conhecimento	Habilidades e conhecimento são desenvolvidos no contexto onde serão utilizados

Tabela 1- Abordagens de aprendizagem tradicional e construtivista
Fonte: REZENDE, 2002

Desse modo, o construtivismo defende a concepção que as pessoas nascem apenas com um conjunto de predisposições neurofisiológicas para a maneira de pensar e agir que

precisam ser desenvolvidos ao longo da vida, propondo que as estruturas mentais devem ser concebidas como o produto de uma construção a partir de prolongadas etapas de reflexão individual e interação com outros indivíduos (ARIAS; YERA, 2012).

De acordo com Fonseca (2009), Seymour Papert (1994), um dos discípulos de Piaget, sugere que uma boa estratégia para ocorrência de aprendizagem é a construção de artefatos, essa concepção é conhecida como Construcionismo.

O construcionismo vai além da afirmação de que o conhecimento é construído pelo indivíduo, mas que isso acontece quando indivíduo está engajado na construção de algo externo, de modo que esse processo nos levaria a um modelo de ciclo contínuo de internacionalização e externalização do que está fora e dentro de nós, respectivamente (REZENDE, 2002).

Seguindo o pensamento de Papert, neste trabalho a pedagogia proposta como metodologia ser seguida no autoaprendizado foi a de “Projetos de Aprendizagem”, uma pedagogia que é fundamentada na construção de artefatos e que explora os princípios do construtivismo e dá suporte ao construcionismo, propondo que o estudante construa o seu conhecimento por meio da exploração de uma questão de investigação. Desse modo, a seguir será feito um detalhamento sobre essa pedagogia.

2.3 PROJETOS DE APRENDIZAGEM

A pedagogia de Projetos de Aprendizagem tem origem nos resultados de pesquisas realizadas pelo Laboratório de Estudos Cognitivos – LEC/UFGRS em “psicologia genética, sobre desenvolvimento da inteligência e sobre o processo de aprendizagem” (SCHELMMER, 2018), sendo um experimento piloto realizado pelo Projeto Amora (CAp/UFGRS) entre os anos de 1995 e 1996. Desde então, essa pedagogia vem sendo adotada em algumas escolas públicas do país a partir de capacitações de professores realizadas por um programa de inserção das tecnologias da informação e da comunicação, denominado PROINFO, que foi desenvolvido pela Secretaria de Educação a Distância do MEC em parceria com os Estados e Municípios, (FAGUNDES et al., 2006).

Essa pedagogia é fundamentada na concepção epistemológica interacionista-construtivista, de modo que torna-se de extrema importância a formulação de questionamentos pelo autor do projeto, ou seja, pelo indivíduo que busca construir seu conhecimento a partir do que já conhece e da sua curiosidade, inquietações, perturbações e conflitos relacionados ao tópico em questão, uma vez que, é a partir desses questionamentos que o estudante irá interagir

com o desconhecido ou com novas situações, com o propósito de se apropriar de conhecimento específico, seja esse conhecimento científico, artístico ou cultural (SCHELMMER, 2018).

De tal modo, o desenvolvimento de um projeto de aprendizagem consiste na busca por informações que esclareçam as indagações do indivíduo sobre sua realidade dentro de um contexto em estudo, sendo assim, Fagundes (et al., 2006) afirma que o objetivo dessa pedagogia é:

[...] o desenvolvimento de um processo de aprendizagem que alcance a construção de novos conhecimentos, em que o aprendiz possa sistematizar informações ampliando sua rede de significações, possa reestruturar o raciocínio lógico sobre os novos significados enquanto elabora sínteses de respostas descritivas e explicativas para sua curiosidade. [...]

Desse modo, enquanto o projeto é desenvolvido, o estudante constrói uma rede de conhecimentos em torno de seu questionamento em investigação. Ainda segundo Fagundes (et al., 2006), a pedagogia de Projetos de Aprendizagem apresenta resultados significativos quanto ao desenvolvimento cognitivo e melhoria da aprendizagem, especialmente na construção de conceitos científicos.

2.3.1 METODOLOGIA DE PROJETOS DE APRENDIZAGEM

A pedagogia de Projetos de Aprendizagem tem por base a construção do conhecimento a partir dos questionamentos sobre o tema em estudo, dessa forma, a metodologia utilizada nessa pedagogia consiste primeiramente em selecionar uma curiosidade, que de acordo com Fagundes (et al. 2016), é denominada didaticamente de “Questão de Investigação”.

A partir da seleção do questionamento é realizado um inventário do conhecimento do indivíduo em relação ao tema em estudo. Esse conhecimento poderá ser categorizado em certezas e dúvidas. De acordo com Fonseca (2009), o conhecimento categorizado como dúvida deverá ser temporário e o processo de investigação deverá buscar saná-lo, enquanto o conhecimento categorizado como certeza é provisório, visto que ele poderá ser validado ou refutado.

Para melhor entendimento de como ocorre o funcionamento dessa metodologia, será exemplificado o passo a passo de como a pedagogia de Projetos de Aprendizagem funciona com base na questão de investigação: “Como funciona a organização de uma colmeia de abelhas?”. Dessa forma, ele expõe as seguintes certezas e dúvidas:

Certezas:

- C1 - Existe uma abelha rainha
- C2 - Existem abelhas operárias e zangões
- C3 - Existe uma hierarquia na colmeia
- C4 - Os zangões são menores que a abelha rainha
- C5 - As abelhas operárias são responsáveis pela produção do mel
- C6 - As abelhas operárias são responsáveis pela polinização de muitas flores ao colherem o pólen para produzir o mel
- C7 - As abelhas com ferrão morrem ao perdê-lo após picar alguém
- C8 - As abelhas sul-americanas não possuem ferrão

Dúvidas:

- D1 - Como é feita a separação das abelhas operárias?
- D2 - Existem separação entre as funções das operárias?
- D3 - Qual a função do zangão para a colméia?
- D4 - Como ocorre a fabricação da cera do mel?
- D5 - Como as abelhas se alimentam?
- D6 - Quantas abelhas existem em uma colméia?
- D7 - Como uma abelha rainha se torna rainha?
- D8 - A rainha pode comandar as outras abelhas?
- D9 - Como é a reação das abelhas em caso de invasão na colmeia?
- D10 - As abelhas atacam sempre que se sentirem ameaçadas?
- D11 - Uma colmeia pode ser mudada de lugar?
- D12 - Já foi comprovado algum nível de inteligência entre as abelhas?
- D13 - Como as abelhas sem ferrão se defendem?
- D14 - Como as abelhas se comunicam?

Após a definição de dúvidas e certezas, vários caminhos podem levar a construção do projeto. No entanto, segundo Fagundes (1998, apud. FONSECA, 2009), nesse processo deve ser respeitada a autonomia para:

- Decidir critérios de julgamento sobre relevância em relação ao tema em estudo;
- Buscar, localizar, selecionar, recolher informações;

- Definir, escolher, inventar procedimentos para testar a relevância das informações escolhidas em relação aos problemas e às questões formuladas;
- Organizar e comunicar o conhecimento construído.

Sendo assim, uma proposta de construção exemplificada por Fonseca é a separação dos conhecimentos elencados em grupos contextuais, como mostrados a seguir:

G1 (tarefas): C5, C6, D2, D3, D4

G2 (hierarquia): C1, C2, C3, D1, D7, D8.

G3 (comunicação): D14

G4 (inteligência): D11, D12

G5 (conflitos): D9, D10, D13.

G6 (população): D6.

G7 (fisiologia): C4, C7, C8, D5.

A coleta de dados é o ponto de partida para construir o conhecimento a partir desse passo em diante. Essa coleta pode ser realizada de diversas maneiras a partir de diversas fontes (livros, entrevistas, revistas e internet). A ideia é que ao realizar a pesquisa e investigação várias dúvidas poderão se tornar certezas; certezas também poderão se tornar dúvidas; além da possibilidade da construção de novas dúvidas e certezas, até que o conhecimento seja construído pelo próprio estudante (SILVA; MENEZES; FAGUNDES, 2016).

Nessa perspectiva, a tecnologia pode ser utilizada como uma ferramenta de auxílio ao autoestudo, de modo a prover acesso a uma gama de conhecimentos atuais sobre determinado tópico em questão, assessorando o estudante na progressão e construção de seu aprendizado ao longo da vida. Daí, surge a necessidade de ambientes que possam auxiliar as pessoas nesse aprendizado autônomo. Esses ambientes são conhecidos como *Personal Learning Environments* – *PLE* ou Ambientes Pessoais de Aprendizagem.

3 AMBIENTES PESSOAIS DE APRENDIZAGEM

Os ambientes pessoais de aprendizagem (*Personal Learning Environments – PLE*) utilizam-se de tecnologias Web atuais e de mídias sociais que auxiliam o estudante individual, ou autônomo, a gerenciar a sua própria aprendizagem (HAWORTH, 2016). De acordo com Motta e Attwell (apud CASTRO et al., 2015), os PLEs são uma abordagem que dão poder de escolha e autonomia ao aprendiz, favorecendo a colaboração e compartilhamento do conhecimento, instrumentando a aprendizagem ao longo da vida.

Os PLEs atribuem ao aluno um papel ativo, colocando-o no centro do processo de aprendizado, ao mesmo tempo que aumenta o controle dos próprios alunos sobre o seu estudo. Assim, com essa assistência, eles podem gerenciar seu próprio ensino e trabalhar em direção a seus objetivos educacionais (Şahin e Uluyol, 2016).

Para Şahin e Uluyol (2016), os PLEs podem ser classificados em determinadas categorias, como por exemplo: enciclopédias, motores de busca, redes sociais, documentos, e-mail, gráficos, blogs, microblogs, chats, apresentações e vídeos.

Haworth (2016) considera que para um ambiente de estudo autodidata, como o proposto neste trabalho, um PLE ideal deveria possuir as seguintes características:

- *fácil de utilizar*: caso seja uma ferramenta difícil de se utilizar, provavelmente as pessoas não irão utilizá-la. Esse conceito serve para administradores, alunos e professores;
- *aberto ('open')*: nesta característica, Haworth considera como aberto um software independente de um Sistema de Gestão de Aprendizagem (SGA), escola, ou programa, isto é, sem um gerenciamento por tutores on-line. Outro sinônimo considerado para este conceito é que o PLE possa ser portátil, ou seja, que o aprendiz possa utilizá-lo em qualquer lugar;
- *dinâmico*: o PLE deve evoluir de acordo com os objetivos e realizações do aluno, evoluindo de acordo com o seu aprendizado;
- *dar opções para colaboração*: segundo com Dabbagh e Kitsantas (2012, apud HAWORTH, 2016), a capacidade de compartilhar informações nos PLEs com outras pessoas tem sido benéfica. No entanto, tal funcionalidade de compartilhamento deve ser opcional de cada estudante.

Existem várias ferramentas bem disseminadas que podem ser utilizadas como PLEs, que se enquadram nas características dadas por Haworth (2016) e nas categorias definidas por Şahin e Uluyol (2016), entre elas, podemos citar algumas:

- *enciclopédias*: wikipédia;
- *motores de busca*: Google, Yahoo, Bing, etc..
- *redes sociais*: Facebook, Instagram;
- *documentos*: Google Docs;
- *e-mail*: Gmail, Outlook;
- *blogs*: Blogger, Wordpress;
- *microblogs*: Twitter, Tumblr;
- *gráficos*: Pinterest, Imgur, Flickr
- *chats*: Hangout, Skype;
- *apresentações*: SlideShare
- *vídeos*: Youtube, Vimeo

Além das que se enquadram nas categorias definidas por Şahin e Uluyol (2016), Haworth (2016) cita ainda como PLEs o website pessoal, social bookmarks como o Reddit e o Symbaloo.Edu.

Embora algumas dessas ferramentas, como o Symbaloo.Edu por exemplo, tenham cunho educacional, nenhuma delas realiza um assessoramento mais profundo ao estudante. Esse assessoramento pode ser realizado por meio de recomendação de informação, que pode ser executada com base dos vestígios de informações deixados pelo usuário, como por exemplo, anotações de estudo.

A ferramenta proposta neste trabalho dispõe-se a realizar esse assessoramento dessa forma. No entanto, para que o sistema entenda esses vestígios é necessário que ele efetue a recuperação da informação passada anteriormente pelo usuário.

4 RECUPERAÇÃO DE INFORMAÇÃO

A recuperação de informação surgiu com o intuito de facilitar a tarefa de se obter informações em um curto espaço de tempo. Técnicas de organização das informações, estratégias de busca e a criação de mecanismo de busca foram criados com o propósito de tornar a recuperação de informação mais rápida e eficaz. De acordo com Mooers (1951, apud FARIAS, 2015) a recuperação de informação trata dos aspectos intelectuais da descrição da informação e sua especificação para busca, bem como qualquer sistema, técnicas ou máquinas que são empregadas para realizar estas operações.

Desse modo, a área de Recuperação de Informação (RI) remete ao armazenamento, organização, representação e acesso a itens de informação, de modo que a representação e a organização dos itens devem prover um fácil acesso as informações de interesse ao usuário.

Um modelo de recuperação de informação pode ter sua especificação formada a partir de três elementos básicos: a representação dos documentos, a representação da expressão de busca e a função de busca. De acordo com (FERNEDA, 2012 apud NETO; FERNEDA, 2016), conforme demonstrado na Figura 2:

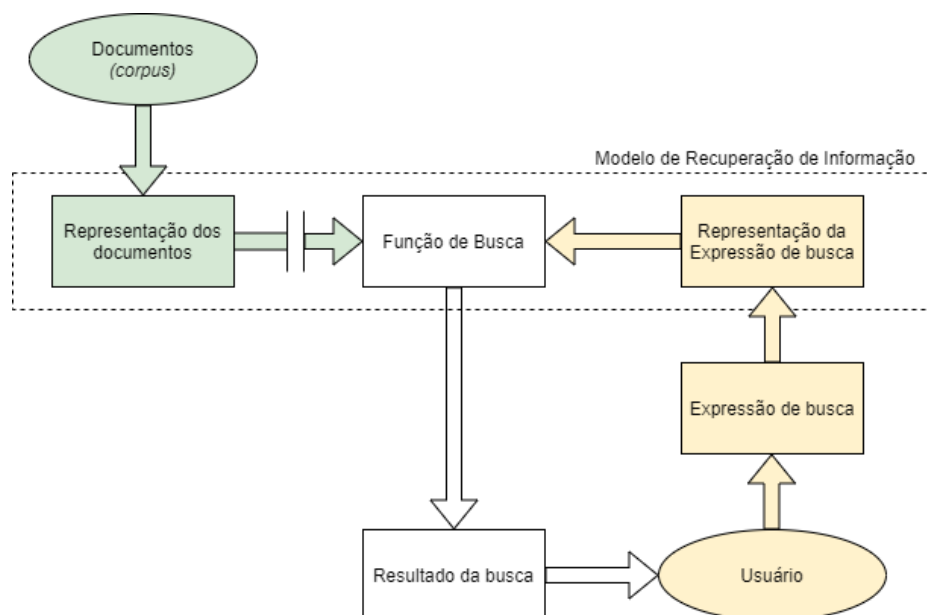


Figura 2 - Representação do processo de recuperação de informação
Fonte: FERNEDA, 2012, apud NETO, 2016

Assim, de acordo com a representação do processo acima, em um sistema de recuperação de informação o “usuário” expressa a sua necessidade de informação por meio de

uma “expressão de busca” composta geralmente por um conjunto de termos. No centro do processo de recuperação de informação está a “função de busca”, que realiza as comparações entre as representações dos documentos e as representações da expressão de busca. Sendo assim, é de extrema importância a forma como a representação dos documentos e os termos de busca são definidos, de forma que a função de busca seja capaz de realizar o cálculo de similaridade entre a expressão de busca e cada um dos documentos do *corpus* com eficácia (FERNEDA, 2012 apud NETO; FERNEDA, 2016).

Portanto, a recuperação efetiva de informações relevantes depende diretamente da linguagem de representação dos itens de informação, isto é, na representação dos documentos de uma coleção por um conjunto de termos indexadores ou por palavras-chave, que podem ser especificadas por um indivíduo ou podem ser extraídas diretamente de um documento de texto. Em contrapartida, ela também depende da forma como os usuários traduzem linguisticamente as suas necessidades de informação (MEADOW et al, 2007, apud FERNEDA; DIAS, 2017).

Por consequência, é fundamental que a expressão de busca utilizada pelos usuários seja representada de forma similar à forma como os documentos foram representados, de modo que seja possível realizar uma comparação entre a representação dos documentos e a necessidade do usuário. Dessa forma, a recuperação efetiva de informações relevantes é diretamente afetada pelas tarefas com as quais o usuário está envolvido (*user tasks*) e pela visão lógica dos documentos adotada pelo sistema de RI (BAEZA-YATES, apud FARIAS, 2015). Na Figura 3 podemos verificar a representação da interação do usuário nas diferentes tarefas identificadas.

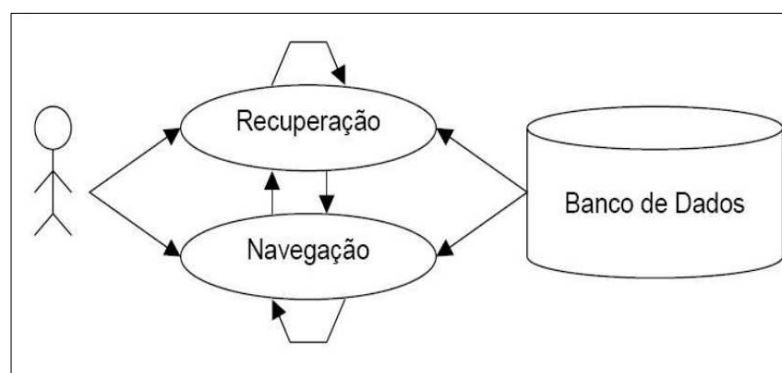


Figura 3 - Interação do usuário com o sistema de recuperação através de diferentes tarefas
Fonte: BAEZA-YATES, apud FARIAS, 2015

Dada a importância sobre a representação dos documentos disponíveis para realizar a recuperação da informação e a variedade de modelos com diferentes funções de busca, a

seguir, será abordado sobre como deve ser a estrutura lógica de um documento e os principais modelos de RI.

4.1 A VISÃO LÓGICA DE UM DOCUMENTO

Por questões históricas e de limitações computacionais, os documentos de uma coleção são frequentemente representados por um conjunto de termos indexadores ou por palavras-chave. Essas palavras-chave podem ser extraídas diretamente de um documento de texto ou podem ser especificadas por um indivíduo. Desconsiderando a origem dessas palavras-chave, quer sejam extraídas automaticamente, quer sejam geradas por um especialista, elas fornecem uma visão lógica do documento.

Os computadores modernos têm possibilitado a representação de documentos por seu conjunto completo de palavras. Nesse caso, Baeza-Yates e Ribeiro-Neto (1999, apud FARIAS, 2015) dizem que o sistema de recuperação utiliza uma visão lógica baseada no texto completo (uma representação). Entretanto, mesmo os computadores mais modernos, no caso de coleções, precisam ser reduzidos o conjunto de palavras-chave representativas; isso pode ser acompanhado da remoção de *stopwords* (como artigos e conectivos), do uso de *stemming* (que substituem palavras flexionadas por seus respectivos radicais), da identificação de substantivos (que eliminam adjetivos, advérbios e verbos) e da compressão das estruturas lógicas internas (índices e índices invertidos, por exemplo). Essas operações, chamadas de operações textuais, reduzem a complexidade da representação dos documentos, além de possibilitarem a conversão de um texto completo em um conjunto de termos indexadores.

Ainda para Baeza-Yates e Ribeiro-Neto (1999, apud FARIAS, 2015), “o documento completo é claramente a mais integral visão lógica de um documento, mas seu uso implica custos computacionais elevados”. Um conjunto resumido de categorias (geradas por um especialista humano) provê a visão lógica mais concisa de um documento, mas seu uso pode levar a uma recuperação de baixa qualidade.

Em alguns casos, visões lógicas intermediárias de um documento podem ser utilizadas por um sistema de RI, conforme ilustrado na Figura 4. Portanto, adotando algumas das representações intermediárias, o sistema de RI pode reconhecer estruturas internas normalmente encontradas em um documento comum, como capítulos, seções, subseções, entre outras (SILVA, 2016).

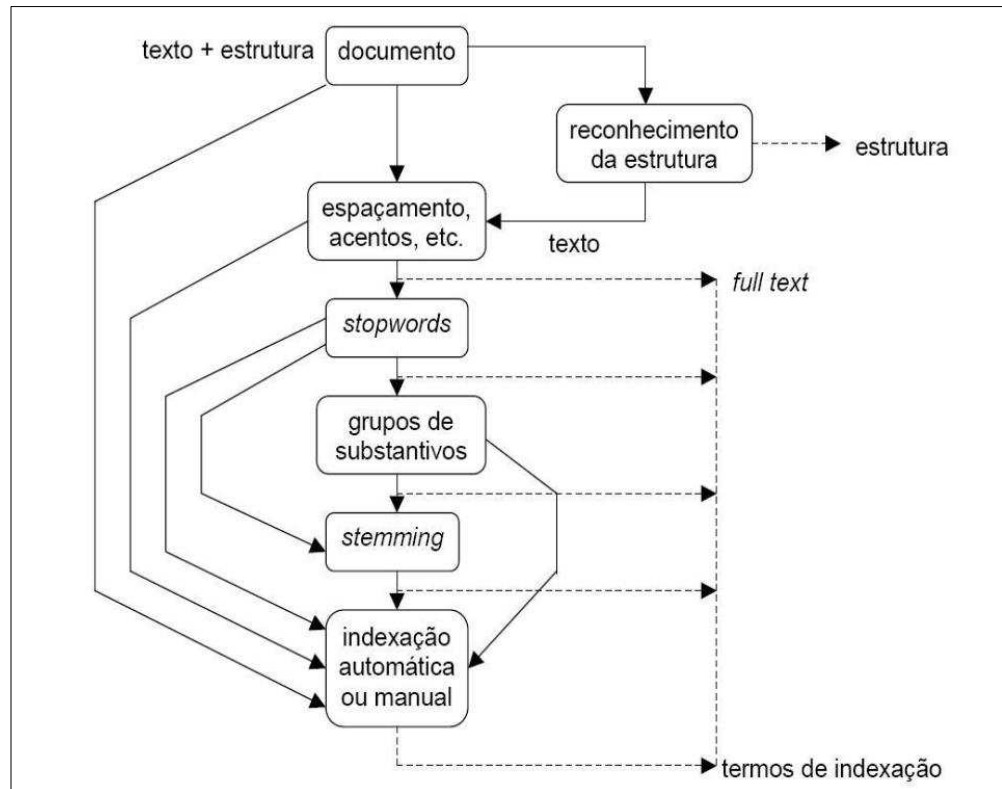


Figura 4 - Visões lógicas intermediárias utilizáveis para sistema de RI
 Fonte: BAEZA-YATES; apud SILVA, 2016

Como mostrado na Figura 4, o problema na representação lógica de um documento é contínuo. Assim, a representação pode variar de um texto completo até uma especificação de alto nível, realizada por um especialista humano. A escolha de uma representação, mais ou menos elaborada, depende da necessidade do sistema de RI a ser implementado. A seguir, serão apresentados os principais modelos de RI.

4.2 MODELOS DE RECUPERAÇÃO DE INFORMAÇÃO

Existem vários modelos de recuperação de informação que são responsáveis por determinar o modo como a função de busca irá se comportar. Para Baeza-Yates e Ribeiro Neto (2013, apud MONTEIRO et al, 2017), os modelos de recuperação de informação podem ser classificados em três categorias: baseados em texto, baseados em links e baseados em objetos multimídia, conforme exposto na Figura 5:

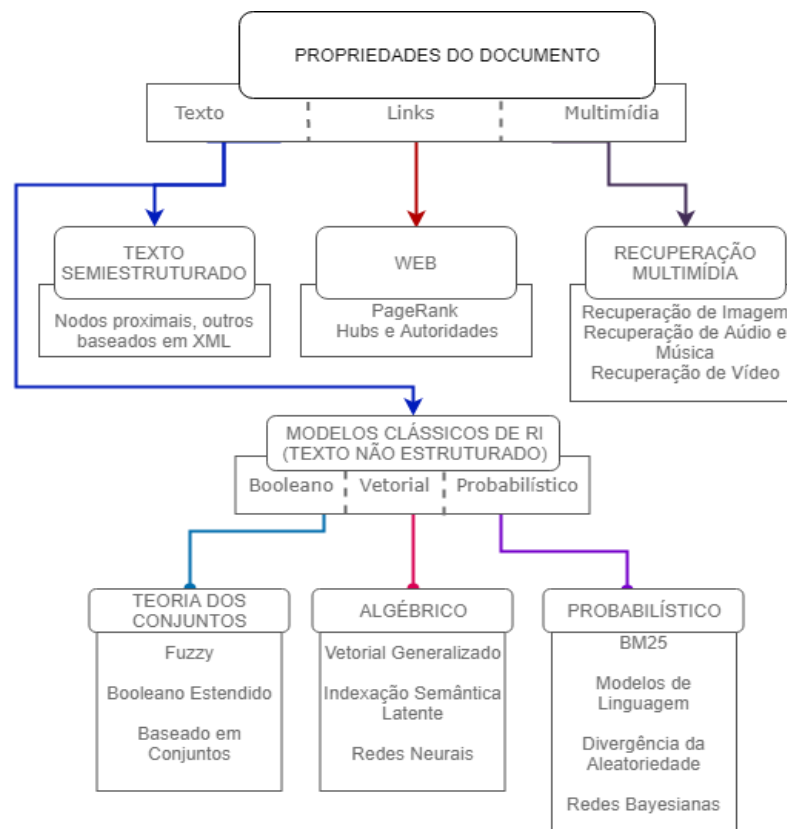


Figura 5 - Taxonomia de modelos de RI
Fonte: Adaptado de MONTEIRO et al (2017)

Na categoria baseada em texto, o texto dos documentos é usado para ranqueá-los de acordo com a consulta. De acordo com Monteiro et al (2017), nos modelos clássicos de RI, booleano, vetorial e probabilístico, o texto é modelado como uma sequência de palavras. Em contrapartida, no texto semiestruturados, os componentes do texto como título, seções e parágrafos fazem parte integral do modelo.

Quanto a evolução dos modelos clássicos, Monteiro et al (2017) afirma que:

A partir dos modelos clássicos, outros foram desenvolvidos visando à ampliação da performance e operacionalidade: baseados na teoria dos conjuntos, a lógica difusa ou nebulosa (*fuzzy*), o booleano estendido e o baseado em conjuntos; baseados nos modelos algébricos, vetorial generalizado, indexação semântica latente e redes neurais; e baseados nos modelos probabilísticos, BM25, modelos de linguagem, divergência da aleatoriedade e redes bayesianas. (MONTEIRO et al, 2017)

Na categoria baseada em links é feito proveito sobre a informação da estrutura, visto que com o grande número de documentos presentes na Web, apenas o ranqueamento baseado

em texto não seria o suficiente. Desse modo, nesta categoria são considerados os links entre as páginas Web (MONTEIRO et al, 2017).

Já na categoria baseada em objetos multimídia, durante a recuperação da informação, o ranqueamento é realizado de forma diferente ou é até inexistente, por causa das particularidades em suas formas e representações (BAEZA-YATES et al, 2013, apud MONTEIRO et al, 2017).

A ferramenta proposta neste trabalho deverá utilizar como expressão de busca as anotações realizadas pelo estudante durante seu autoestudo, ou seja, um texto não estruturado, que precisará ser pré-processado a fim de retirar os termos realmente relevantes e entender a necessidade do usuário para que os documentos recuperados e que representem melhor a sua necessidade, possam ser recomendados. Sendo assim, cada um dos modelos clássicos de RI (booleano, vetorial e probabilístico) mostrados na imagem serão abordados a seguir.

4.2.1 MODELO BOOLEANO

O modelo booleano é o modelo de recuperação simples, baseado na teoria de conjunto e na álgebra booleana. Nesse modelo, um documento é representado por um conjunto de termos indexados. As buscas são realizadas através de expressões booleanas, compostas por termos ligados através dos conectivos lógicos *and*, *or* e *not* (e, ou e não) e, como resultado, apresenta-se os documentos que satisfaçam às restrições lógicas da expressão de busca.

De acordo com Fonseca (2009), as principais vantagens desse modelo são:

- a expressividade completa, caso o usuário souber exatamente o que quer;
- a facilidade da programação exata da consulta desejada.

No entanto, como este modelo trabalha com lógica binária, ou o documento é relevante ou não é, e não existe uma forma de ordenar os resultados que satisfaçam a consulta do usuário. Sendo assim, este modelo tem como desvantagem:

- o usuário ter que lidar com conhecimento parcial;
- a possibilidade da saída ser nula ou de haver sobrecarga;
- a saída não ordenada por relevância;

Apesar dessas limitações, este modelo ainda é altamente utilizado em sistemas comerciais. Em contrapartida, existem outros modelos que já provêm soluções para as desvantagens abordadas.

4.2.2 MODELO VETORIAL

O modelo vetorial representa um documento por um vetor de termos, onde cada termo possui um valor de importância, ou peso, associado a ele. Do mesmo modo, a função de busca também é representada por um vetor numérico em que cada elemento representa a relevância (peso) de cada termo na representação da necessidade de informação do usuário (SALTON, et al., 1975, apud NETO; FERNEDA, 2016).

Assim, esses pesos são utilizados para medir o grau de similaridade entre cada documento armazenado e a consulta do usuário, de modo que seja possível mensurar uma similaridade parcial, devido a atribuição de pesos não binários para os documentos, havendo uma vantagem em relação ao modelo booleano nesse sentido.

No modelo vetorial, cada documento é representado como um vetor de termos. Cada termo possui um valor associado, que indica o grau de importância (peso - weight) deste no documento. Em outras palavras, cada documento possui um vetor associado, constituído por pares de elementos na forma {(palavra_1, peso_1), (palavra_2, peso_2),..., (palavra_n, peso_n)} (FARIAS, 2015).

Os documentos são ordenados pelo grau de similaridade de forma decrescente, e este modelo provê uma melhor resposta que se ajusta mais a necessidade do usuário do que a resposta que o modelo booleano gera (SILVA, 2016). Essa organização dos documentos pode ser realizada em um arquivo invertido, como demonstrado na Figura 6:

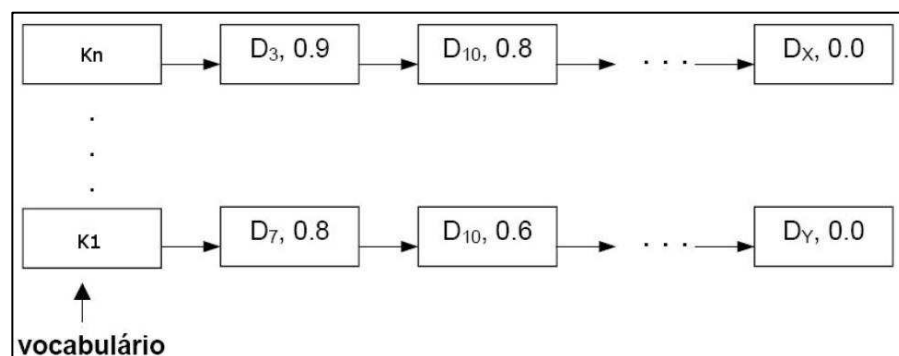


Figura 6 - Documentos Indexados com os pesos de cada documento.
Fonte: YATES, 1999 apud FARIAS, 2015

A utilização de uma mesma representação, tanto para os documentos como para as expressões de busca, permite calcular o grau de similaridade entre o vetor que representa uma determinada expressão de busca e cada um dos vetores que representa os documentos. Dessa

forma, em um espaço vetorial contendo N dimensões, a similaridade (*sim*) entre um documento d_j e uma consulta q pode ser obtida a partir da seguinte fórmula (SALTON, 1983 apud NETO; FERNEDA, 2016):

$$sim(d_j, q) = \frac{\sum_{i=1}^N (w_{i,j} \cdot w_{i,q})}{\sqrt{\sum_{i=1}^N w_{i,j}^2} \cdot \sqrt{\sum_{i=1}^N w_{i,q}^2}}$$

Equação 1 - Cálculo de similaridade no modelo booleano
Fonte: SALTON, 1983 apud NETO; FERNEDA, 2016

Onde $w_{i,j}$ é o peso do i-ésimo termo do documento d_j e $w_{i,q}$ é o peso do i-ésimo termo da expressão de busca q .

O cálculo da similaridade (*sim*) é realizado entre o vetor que representa a expressão de busca do usuário e cada um dos vetores representativos de cada documento na coleção de dados. Os valores da similaridade (*sim*) entre a expressão de busca e cada um dos documentos são utilizados no ordenamento do resultado de busca (NETO; FERNEDA, 2016). Todo esse processo pode ser melhor entendido a partir da Figura 7:

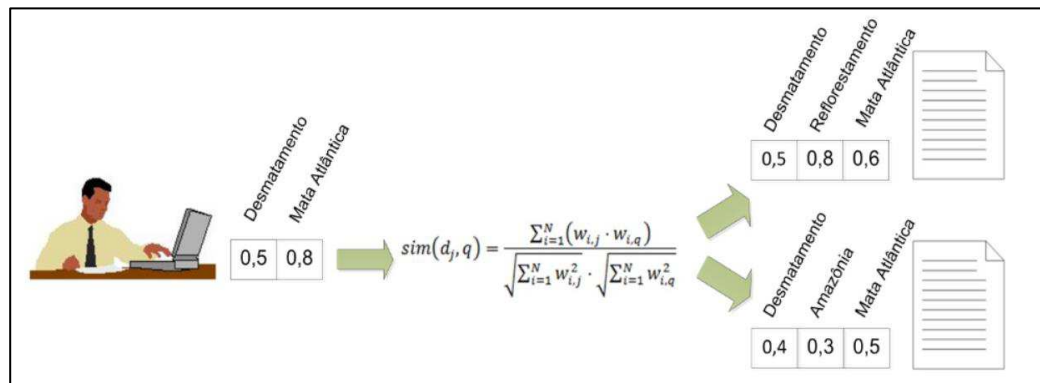


Figura 7- Ilustração do Modelo Vetorial
Fonte: NETO; FERNEDA, 2016

Apesar de sua simplicidade, o modelo vetorial é uma estratégia de ranque bastante eficaz. Sendo assim, de acordo com Farias (2015) o modelo vetorial apresenta como principais vantagens:

- Seu esquema de atribuição de pesos, que melhora a qualidade da recuperação;
- Sua estratégia de correspondência parcial, que possibilita a recuperação de documentos que se aproximam da consulta;

- Sua fórmula de ranque, baseada em cosseno, que classifica os documentos de acordo com seu grau de similaridade.

4.2.3 MODELO PROBABILÍSTICO

O modelo probabilístico aborda a Recuperação de Informações como um problema estatístico. Ele se fundamenta na ideia que a partir de uma consulta do usuário, o conjunto de documentos da coleção é subdividido em quatro grupos (FARIAS, 2015):

- o conjunto de documentos relevantes (Rel);
- o conjunto dos documentos recuperados (Rec);
- o conjunto dos documentos relevantes recuperados (RR); e
- o conjunto dos documentos não relevantes e não recuperados.

O que se busca, nesse modelo, é saber a probabilidade de um documento D ser ou não relevante para uma consulta. Tal informação pode ser obtida assumindo-se que a distribuição de termos, na coleção, seja capaz de informar a provável relevância para um dado documento (SILVA, 2016).

Do mesmo modo que no modelo vetorial, no modelo probabilístico, uma consulta é um subconjunto de termos de indexação e o documento é representado por um vetor de pesos que definem a existência ou não de termos de indexação. Assim, o objetivo é tentar estimar a probabilidade de um documento ser relevante de acordo com uma consulta q do usuário, considerando as informações disponíveis ao sistema (CRUZ, 2019).

Dessa forma, pode-se pensar no processo de busca como um procedimento de especificação das propriedades de um conjunto resposta ideal. No entanto, descobrir exatamente o que são essas propriedades torna-se um problema, uma vez que tudo o que sabemos sobre elas é que são termos indexados dos quais a semântica pode ser utilizada para caracterizá-las (FARIAS, 2015).

Consoante com Silva (2016), o modelo probabilístico é baseado no seguinte pressuposto: “dada uma consulta q e um documento d_j em uma coleção, o modelo probabilístico tenta estimar a probabilidade do usuário considerar o documento d_j interessante (i.e. relevante)”. No entanto, o modelo assume que essa possibilidade de relevância depende apenas da representação da consulta do documento.

Além do mais, o modelo também assume que existe um subconjunto de documentos que o usuário prefere definir como resposta a sua consulta q , esse conjunto é dado como

conjunto de resposta ideal e é chamado de R . O conjunto R deve maximizar a probabilidade global de relevância para o usuário. Os documentos pertencentes a esse conjunto R são classificados como “relevantes para a consulta”, enquanto os demais documentos são classificados como “não relevantes” (SILVA, 2016).

Sendo assim, dada uma consulta q , o modelo probabilístico associa a cada documento d_j , como uma medida de sua similaridade a consulta, a razão $P(d_j \text{ relevante para } q)/P(d_j \text{ não relevante para } q)$ que computa as probabilidades do documento d_j ser relevante para a consulta q . Tomando as probabilidades de relevância como o ranque, minimiza-se a probabilidade de um julgamento errôneo (FARIAS, 2015).

Dessa forma, o modelo probabilístico tem como ponto forte o ranqueamento dos documentos de maneira decrescente, conforme a probabilidade de relevância. Apesar disso, o método tem como pontos fracos a não consideração da frequência de ocorrência do termo de indexação e a necessidade de estimar a separação dos documentos em relevantes e irrelevantes (BARTH, 2013 apud CRUZ, 2019).

5 SISTEMAS DE RECOMENDAÇÃO

Mesmo com uma recuperação de informação satisfatória, ainda há o perigo do retorno de algo irrelevante para o usuário. Este problema acaba causando um desestímulo, pois há dificuldade de encontrar o que se realmente deseja. Como consequência, o usuário se sente perdido em meio a um mundo de informações sem saber como encontrar o que lhe interessa.

Conforme Fonseca, Farias e Silva (2014), os sistemas de recomendação (SR) surgiram como uma resposta à dificuldade das pessoas de escolher em meio a grande variedade de produtos e serviços e as várias alternativas apresentadas. Desse modo, eles pretendem auxiliar no processo de indicar e receber indicações, de forma a facilitar a busca por conteúdo interessante ao usuário prevendo itens que possam ser relevantes ao mesmo.

De acordo com Farias (2015), sistemas de recomendação são ferramentas de software e técnicas que fornecem sugestões de itens que sejam úteis para um usuário. Para Ricci et al. (2011, apud FARIAS, 2015):

Os Sistemas de Recomendação tentam prever quais são os produtos ou serviços mais adequados, com base nas preferências e restrições do usuário. Dessa forma, os sistemas de recomendação coletam dos usuários suas preferências, que podem ser explícitas, através das avaliações de produtos, ou implícitas, inferidos através da interpretação das ações do usuário.

Sendo assim, a recomendação de informação pode ser vista como uma forma particular de filtragem de informação por meio dos comportamentos e perfil do usuário, gerando como resultado uma lista de itens que deverá ser associada as suas preferências. Desta forma, os sistemas de recomendação permitem a personalização de resultados de forma proativa, sem que o usuário precise formular uma expressão de busca para encontrar o que ele necessita (KREBS; ROCHA; RIBEIRO, 2017).

Portanto, os sistemas de recomendação envolvem a construção de um modelo ou perfil de interesses do usuário. Segundo Fonseca, Farias e Silva (2014), a construção deste perfil varia de acordo com a técnica de recomendação utilizada tendo como as principais técnicas de recomendação a Filtragem Baseada em Conteúdo, Filtragem Colaborativa e o Modelo Híbrido. A seguir serão detalhados esses três modelos.

5.1 FILTRAGEM BASEADA EM CONTEÚDO

A filtragem baseada em conteúdo parte do princípio de que os usuários tendem a se interessar por itens similares aos que lhe tenham interessado no passado. Esta abordagem tem como objetivo gerar descrições dos conteúdos dos itens e compará-los com os interesses do usuário para verificar a sua relevância.

Para Ricci et al. (2011, apud FARIAS, 2015),

[...] o processo básico realizado por uma recomendação baseada em conteúdo consiste em combinar os atributos de um perfil de usuário em que as preferências e interesses são armazenados, com os atributos de um objeto de conteúdo (item), a fim de recomendar ao usuário novos itens.

Este modelo de filtragem possui suas origens na área de recuperação e filtragem de informação, tendo assim um enfoque na recomendação de itens com informações textuais, tais como documentos e sites. O conteúdo nesse tipo de filtragem é geralmente descrito por palavras chave (BARBOSA, 2014). O cálculo da importância dessas palavras em um documento pode ser definido de diferentes maneiras, sendo a mais comum o modelo estatístico $TF*IDF$ ¹ proposto por Salton e Buckley (1988, apud FARIAS, 2015).

A aplicação de técnicas de indexação permite mensurar a relevância de um termo para o documento, e então é feita a comparação do perfil do usuário para verificar a similaridade, ou seja, se é relevante a ele. De acordo com Farias (2015), o perfil do usuário pode ser obtido de duas formas:

- Explicitamente: ocorre quando o usuário fornece informações sobre o seu perfil de forma direta por meio de questionários, e a partir das respostas, o sistema pode obter os seus interesses;
- Implicitamente: ocorre quando o sistema verifica automaticamente as preferências do usuário de acordo com o seu comportamento no sistema.

Uma desvantagem dessa abordagem é que o sistema tende a recomendar itens que já foram avaliados pelo usuário, pois o sistema irá recomendar apenas com base em itens já visualizados pelo usuário, enquanto, segundo Medeiros (2013), a novidade no item recomendado é uma característica desejável em sistemas de recomendação.

¹ Modelo de ponderação que atribui um peso maior para palavras que aparecem muito em um determinado documento e em poucos documentos da base de dados. (BARBOSA, 2014)

5.2 FILTRAGEM COLABORATIVA

A Filtragem Colaborativa surge da tentativa de imitar o comportamento de um usuário pela solicitação de recomendação a outras pessoas. De tal modo, o usuário avalia itens e de acordo com as avaliações dadas, são criados grupos de usuários que apresentam perfis semelhantes. Dessa forma, caso um grupo de usuários tenha avaliado com uma boa nota um certo item que um outro usuário ainda não tenha avaliado, pressupõe-se que esse usuário achará esse item relevante.

Segundo Jannach (2010, apud FARIAS, 2015), “a ideia básica destes sistemas é que, se os usuários compartilhavam os mesmos interesses no passado [...] eles também têm gostos semelhantes no futuro”. Dessa forma, presume-se que caso o usuário A e o usuário B possuem um histórico de compras similares e o usuário B comprou um livro que o usuário A ainda não tenha visualizado, a lógica básica é recomendar esse mesmo livro ao usuário A.

De acordo com Schafer et al. (2009, apud CAZELLA; DRUMM; BARBOSA, 2009), a técnica de recomendação colaborativa constitui-se em uma das técnicas de recomendação mais populares, sendo utilizada em muitos sistemas existentes na Internet.

Geralmente nesse tipo de filtragem existe uma lista de m , usuários $U = \{u_1, u_2, \dots, u_m\}$ e uma lista de n itens $I = \{i_1, i_2, \dots, i_n\}$, onde cada usuário (u_i) tenha uma lista de itens (I_{u_i}) já avaliados pelo mesmo. Também é definido que existe um usuário alvo (u_a) na lista de usuários para o qual a tarefa de recomendação é realizada (MEDEIROS, 2013).

Os algoritmos de recomendação de informação representam a relação usuário-item como uma matriz $m \times n$ de avaliações (A), onde cada entrada $A_{i,j}$ representa a avaliação do i -ésimo usuário i para o j -ésimo item j (MEDEIROS, 2013).

Segundo KONSTAN (2004, apud BARBOSA, 2014), as técnicas de filtragem colaborativa ocorrem em três etapas:

- 1ª etapa:** a formação da vizinhança de usuários para o usuário alvo de acordo com o valor obtido no cálculo de similaridade;
- 2ª etapa:** cálculo da previsão de notas entre um usuário e os itens que ele ainda não avaliou de acordo com o resultado da similaridade;
- 3ª etapa:** recomendação dos itens mais bem avaliados.

Existem duas técnicas de filtragem colaborativa: a baseada em memória e a baseada em modelo. A filtragem colaborativa baseada em memória utiliza toda a matriz usuário x item para realização do cálculo de similaridade e previsões, enquanto que a filtragem colaborativa

baseada em modelo usam um conjunto de avaliações para aprender um modelo, restringido apenas as avaliações presentes nesse grupo de dados e não na base inteira, como realizado na filtragem colaborativa baseada em memória (BARBOSA, 2014).

5.3 MODELO HÍBRIDO

O modelo de filtragem híbrida tem por objetivo a superação das limitações individuais da filtragem baseada em conteúdo e filtragem colaborativa. Dessa forma, ela promove uma combinação entre os pontos fortes dessas duas para criar técnicas que podem ter um desempenho mais robusto em uma ampla variedade de configurações (AGGARWAL, 2016).

6 DESCRIÇÃO DO SISTEMA

A ferramenta proposta neste trabalho, tem como objetivo o armazenamento de informações realizadas durante o autoestudo e a possibilidade de recuperar e recomendar informações de acordo com o andamento do estudo do usuário e a possibilidade de conexão do sistema proposto com redes sociais.

Para alcançar esse objetivo, deve-se analisar e escolher as melhores técnicas de recuperação e recomendação de informação para a necessidade proposta. Além do mais, é necessário definir quais tecnologias atuais se adequam melhor à construção do protótipo, bem como definir qual a melhor arquitetura e forma de comunicação entre seus componentes.

Assim, de forma a obter um melhor entendimento, cada definição do projeto realizada até o momento será percorrida em subtópicos, iniciando pela definição do método de recuperação de informação.

6.1 DEFINIÇÃO DO MÉTODO DE RECUPERAÇÃO DE INFORMAÇÃO

A ferramenta proposta neste trabalho deverá utilizar como expressão de busca as anotações realizadas pelo estudante durante seu autoestudo, ou seja, um texto não estruturado, que precisará ser pré-processado a fim de retirar os termos realmente relevantes e entender a necessidade do usuário para que os documentos recuperados e que representem melhor a sua necessidade, possam ser recomendados.

Desse modo, é necessário a filtragem do conjunto de palavras-chave representativas retiradas das anotações para uma melhor recuperação. Segundo Silva (2016):

[...] isso pode ser acompanhado da remoção de stopwords (como artigos e conectivos), o uso de stemming (que substituem flexionadas por seus respectivos radicais), da identificação de substantivos (que eliminam adjetivos, advérbios e verbos) e da compressão das estruturas lógicas internas (índices e índices invertidos, por exemplo). Essas operações, chamadas de *operações textuais*, reduzem a complexidade da representação dos documentos, além de possibilitarem a conversão de um texto completo em um conjunto de termos indexadores. (SILVA, 2016)

Sendo assim, será realizado pré-processamento das dúvidas do usuário, e realizada a ponderação dos termos extraídos. De acordo com Farias (2015), “a ponderação de termos é

uma importante ferramenta para determinar a relevância de uma palavra, tanto em relação ao documento que está contida, quanto em relação à coleção de documentos”.

A principal técnica utilizada para realizar a ponderação de termos é a $TF \cdot IDF$:

$$TF(t_i, d_j) = \frac{f_{i,j}}{\max_i f_{i,j}}$$

Equação 2 - Cálculo TF
Fonte: Farias, 2015

$$IDF = \log\left(\frac{N}{n_i}\right)$$

Equação 3 - Cálculo IDF
Fonte: Farias, 2015

Onde:

$f_{i,j}$ – frequência do termo i no documento d_j ;

$\max_i f_{i,j}$ – máximo sobre as frequências;

N – número de documentos da coleção;

n_i – quantidade de documentos em que o termo ocorre pelo menos uma vez.

O método $TF \cdot IDF$ (*Term Frequency x Inverse Document Frequency*) tem por principal ideia atribuir um peso maior para palavras que aparecem muito em um determinado documento e em poucos documentos da base de dados (BARBOSA, 2014).

Segundo Farias (2015), esse método é mais utilizado para realizar a busca de termos, uma vez que tem por principal objetivo encontrar um documento relevante no meio de vários, podendo não levar em consideração termos frequentes em vários documentos, uma vez que ele tende a dar um peso mais elevado a termos que aparecem em apenas um documento.

Para o cálculo de similaridade será utilizado uma das técnicas dos modelos clássicos, visto que nesta ferramenta busca-se a recuperação de informação de um texto não estruturado. Assim, será utilizado o modelo vetorial. No modelo vetorial, um documento é representado por um vetor numérico onde cada elemento representa a relevância, ou peso, do respectivo termo no documento.

Desse modo, a recuperação de informação ocorrerá da seguinte forma:

1. Será realizado o pré-processamento e ponderação dos termos encontrados em cada dúvida do usuário, sendo cada dúvida considerado o documento diferente;
2. A dúvida que possuir a menor ponderação será submetida ao motor de busca pela aplicação (Google e Bing);
3. Os títulos das páginas retornadas também serão pré-processados e ponderados;
4. Será realizado o cálculo de similaridade do modelo vetorial conforme o peso de cada termo;
5. As cinco primeiras urls que apresentarem maior similaridade serão retornadas como recomendação na página do estudante;

6.2 DEFINIÇÃO DO SISTEMA DE RECOMENDAÇÃO

Os modelos de filtragem baseada em conteúdo e filtragem colaborativa são técnicas para recomendação de informação, das quais a baseada em conteúdo prioriza o usuário alvo e suas ações no sistema como o único parâmetro necessário para realização da recomendação, enquanto que a filtragem colaborativa depende das ações do usuário alvo e de usuários de perfis semelhantes a ele para realização da recomendação da informação, desse modo, a filtragem colaborativa consegue recomendar itens de assuntos nunca buscados pelo usuário alvo.

No entanto, deve-se frisar que em ambas as técnicas têm suas vantagens e desvantagens em relação a outra. Sendo assim, enquanto na filtragem baseada em conteúdo pode ocorrer do sistema recomendar itens que já foram avaliados pelo usuário, visto que o SR utiliza como base apenas itens já visualizados pelo mesmo, na filtragem colaborativa existe a dependência da construção de uma matriz de usuários similares, para que possa realizar sua recomendação.

Contudo, por conta de suas características, a recomendação baseada em conteúdo consegue realizar uma recomendação mais precisa em sua primeira recomendação, já que depende apenas do usuário final, ao mesmo tempo que a filtragem colaborativa consegue realizar a recomendação de itens nunca buscados pelo usuário, mas que podem ser de seu interesse. Além desses modelos, ainda temos o modelo híbrido que se utiliza das vantagens entre esses dois modelos, a partir da combinação de ambos.

Porém, os modelos descritos até o momento consideram apenas duas dimensões para recomendar: os usuários do sistema e os itens disponíveis, não considerando o contexto em que os usuários estão inseridos (DE PAULA A. TRAMONTIN et al, 2018).

Existe várias definições de dimensões que fazem parte do contexto, tais como:

[...] localização (e.g., usuário em casa ou na universidade), tempo (e.g., se é dia de semana ou final de semana, quanto tempo o usuário tem disponível), condições físicas (e.g., barulho), infraestrutura (e.g., dispositivo, velocidade da internet), usuário (e.g., perfil, preferências), tarefa, ambiente social (e.g., outros usuários perto do usuário analisado). (DE PAULA A. TRAMONTIN et al, 2018)

Nas redes sociais, observa-se que o contexto de um determinado usuário depende fortemente de seus conhecidos, de modo que a preferência individual e a influência coletiva são fatores contextuais importantes para as recomendações sociais, visto que estas afetam as decisões dos usuários sobre a retenção de informações.

Sendo assim, este sistema propõem um modelo híbrido que também considera o contexto social em que o usuário está inserido. Deste modo, o protótipo deverá possuir comunicação por redes sociais, por meio do login via Google, de modo que o usuário possa escolher realizar o login por seu intermédio e inclusive realizar a recomendação de informação direta entre seus contatos. Enquanto a filtragem colaborativa poderá utilizar-se dos “amigos” do usuário nessa rede para determinação de sua matriz de similaridade.

Desta maneira, cada modelo de filtragem deverá ter um peso atribuído, sendo realizada a recomendação na forma de filas de prioridade conforme mostrado na Figura 8:

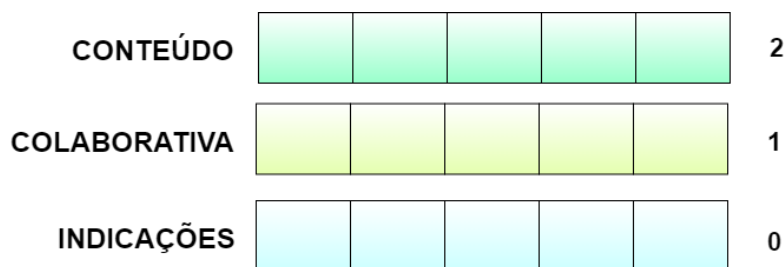


Figura 8 - Filas de prioridade dos modelos de recomendação
Fonte: Autor

Ou seja, a ideia é que inicialmente são propostas as recomendações por indicações de amigos das redes sociais, conseguinte as recomendações por filtragem colaborativa e por fim as recomendações por conteúdo.

No entanto, não foi possível criar um método para verificação de similaridade entre os usuários do protótipo. Como o sistema busca a recomendação dentro do contexto em que está inserido, foi realizada apenas a filtragem via conteúdo e por indicações de contatos. Desta maneira, a recomendação na forma de filas de prioridade do protótipo ficou conforme mostrado na Figura 9:

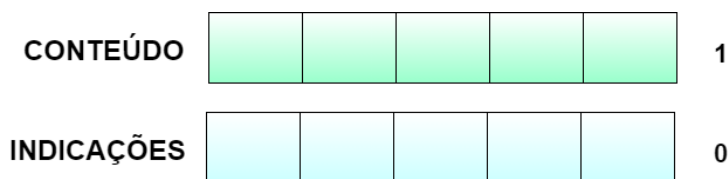


Figura 9 – Filas de prioridade dos modelos de recomendação do protótipo
Fonte: Autor

Ou seja, inicialmente serão propostas as recomendações por indicações de amigos das redes sociais, conseguinte pelas recomendações por conteúdo.

A utilização de uma técnica de recomendação de informação exige a capacidade de processamento de uma grande quantidade de dados. Essa capacidade exigida pode ser encontrada em bancos de dados NoSQL (*Not Only SQL*). A seguir, será realizada uma breve explanação das tecnologias utilizadas para desenvolvimento do protótipo proposto.

6.3 DEFINIÇÃO DE TECNOLOGIAS

Nesta subseção são apresentadas as tecnologias que foram empregadas para desenvolver este trabalho, assim como a linguagem de programação e uma breve visão de como cada tecnologia funciona.

6.2.1 BANCO DE DADOS

A utilização de uma técnica de recomendação de informação exige a capacidade de processamento de uma grande quantidade de dados. Essa capacidade exigida pode ser encontrada em bancos de dados NoSQL. Dessa forma, na escolha de banco de dados (BD), foi escolhido um BD não relacional (NoSQL – *Not Only SQL*).

Os bancos de dados NoSQL surgiram da necessidade de uma melhor performance e de alta escalabilidade em relação aos bancos de dados relacionais. Os BDs relacionais têm sido utilizados como Sistemas Gerenciadores de Banco de Dados (SGBDs) há muito tempo em larga escala em todo o mundo. Porém, com o aumento do volume de dados, que tende a crescer a cada momento, fatores limitantes começaram a ser observados nesse paradigma. (KOKAY, 2012).

De acordo com Saladage e Fowler (2019) cada solução NoSQL possui um modelo diferente que pode ser dividido em quatro categorias vastamente utilizadas: chave-valor, documento, famílias de colunas e grafos. Dessas categorias de armazenamento de bancos de dados NoSQL citados, o modelo orientado a documentos facilita a descrição de metadados de objetos, uma vez que são dispostos no formato chave-valor, dispondo ou não de mais de um atributo, além do fato que aceitam vários tipos de dados.

Segundo Kokay (2012), os bancos de dados mais comuns desse tipo são o MongoDB e o CouchDB. O MongoDB é um caso de sucesso desse tipo de armazenamento, sendo utilizado por grandes empresas tais como a Adobe, Forbes, EA Sports, Ebay, Cisco e Pearson, uma grande líder na área da educação global online (MONGODB, 2020). Sendo assim, ele foi escolhido como o banco de dados na construção o protótipo proposto.

6.2.1.1 MONGODB

O MongoDB é um banco de dados *Not Only SQL* (NoSQL), *open-source* (código aberto), projetado para um fácil escalonamento, *schemaless* e orientado a documentos JSON (JavaScript Object Notation). Por ser *schemaless*, o MongoDB suporta a não existência de um padrão de dados definido, criando uma flexibilidade para definição de requisitos. Deve-se frisar que esta propriedade se torna de grande importância no desenvolvimento deste trabalho, visto a volatilidade de requisitos que se pode esperar no tipo de ferramenta proposta.

A forma de armazenamento no MongoDB é por meio de documentos JSON, esses documentos são inseridos em coleções, que por fim estão inseridas no banco de dados (MONGODB, 2020). Essa hierarquia pode ser melhor entendida conforme a Figura 10 abaixo:



Figura 10 - Estrutura do Banco de Dados MongoDB

Fonte: Autor

Não há limite da quantidade de coleções ou documentos em um banco de dados, porém há limite na quantidade de dados que um documento pode possuir (16 MB). Esses documentos são compostos por *'fields'* e *'values'*, do qual os campos (*fields*) são preenchidos com o seu valor (*value*). O MongoDB é *schemaless*, ou seja, os documentos em uma coleção não precisam estar montados em um mesmo padrão, de modo que não seja necessário que todos os documentos em uma coleção possuam exatamente os mesmos campos, como ocorre em bancos de dados relacionais. (MONGODB, 2020)

Por possuir documentos em formato JSON, o MongoDB torna-se mais próximo das linguagens de programação, uma vez que suporta todos os tipos de dados suportados em documentos BSON (representação binária do JSON). A documentação do MongoDB (2020) cita os seguintes tipos de dados BSON suportados pelo banco:

- String;
- Array;
- Integer;
- Double;
- Boolean;
- Objects;

Além dos tipos especiais do próprio MongoDB:

- ObjectID;
- BinaryData;
- Regular Expression;
- Code;

Ele também pode suportar outros documentos dentro dele. Dessa forma, um exemplo de documento JSON com as abordagens citadas no texto, pode ser observado na Tabela 2 abaixo:

Documento JSON 1	<pre>{ nome: "Maria", idade: 36 endereço: { rua: "Rua dos Alfeneiros", bairro: "Santa Edwirges", numero: 404 } }</pre>
Documento JSON 2	<pre>{ nome: "João", idade: 24, frutas_favoritas: ["manga", "uva", "mamão", "laranja"] }</pre>

Tabela 2 – Exemplificação de documentos JSON

Fonte: Autor

Nessa tabela podemos verificar dois documentos JSON, o primeiro possui como campos *nome*, *idade* e *endereço*; sendo cada campo preenchido com informações do tipo *string*, *integer* e *document*; já no segundo temos como campos *nome*, *idade* e *cores favoritas*; sendo cada campo preenchido com informações do tipo *string*, *integer* e *array*. Desse modo, percebemos a aplicação do *schemaless* e a variedade de tipos que o MongoDB suporta em vários documentos.

Esses documentos ficam armazenados em coleções. Uma coleção é formada por um grupo de documentos que compartilham um conjunto de índices em comum (MONGODB, 2020). Elas são análogas as tabelas em um banco de dados relacional.

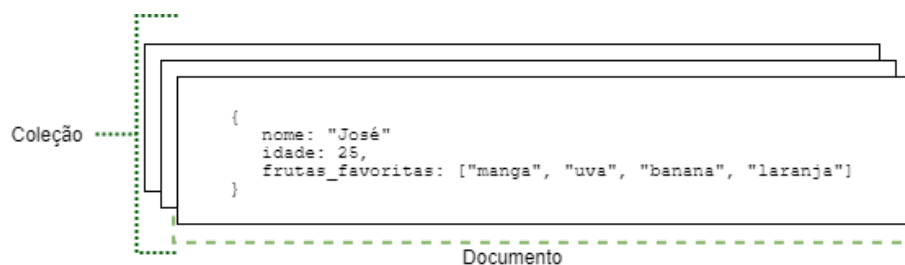


Figura 11 – Coleção

Fonte: Autor

Uma consulta em um banco de dados MongoDB tem como alvo a recuperação de um conjunto específico de documentos. Essa especificação de critérios, ou condições, proposta durante a consulta, identifica os documentos requeridos ao banco pelo cliente.

6.2.1.2 MODELO DE DADOS

Após a escolha do banco de dados para desenvolvimento da ferramenta, partiu-se para o modelo de dados, ou seja, a representação como os dados devem ser organizados no protótipo proposto. Como o banco de dados escolhido foi um banco não relacional, o seguinte formato de documentos JSON foi escolhido:

```
{
  "usuario":{
    _id: 1,
    nome: "Leilanne",
    projetos_de_aprendizagem: [
      {
        questionamento: "Como funciona a organização de um
formigueiro?",
        duvidas: [
          "Qual a quantidade de formigas em um formigueiro? ",
          "Como se escolhe a rainha de um formigueiro? ",
          "Existe apenas uma rainha no formigueiro? ",
          "Como as formigas se comunicam umas com as outras?
",
          "A rainha pode comandar as outras formigas, tomar
decisões? ",
          "Já foi comprovado algum nível de inteligência em um
formigueiro? ",
          "Que atividades inteligentes já foram identificadas
em formigueiros? ",
        ],
        certezas:[
          "Existe uma subdivisão de tarefas no formigueiro",
          "As formigas trabalham em função de proteger a
rainha",
```

```

        "As formigas responsáveis por buscar comida e manter
o formigueiro são as operárias",
        "As formigas podem se comunicar umas com as outras",
        "Existe uma hierarquia no formigueiro",
        "O formigueiro tem compartimentos com funções bem
definidas",
        "Um dos compartimentos do formigueiro é o depósito
de lixo"
    ],

    recomendacoes_aceitas:["https://www.infoescola.com/insetos/fo
rmigueiro/",
    "https://www.portaldosanimais.com.br/informacoes/tudo-sobre-
formigas-e-o-formigueiro/"
    ],
    recomendacoes_recusadas:[
        "https://exame.com/ciencia/formigueiros-na-ilha-
do-cardoso-tem-ate-12-rainhas-por-colon/",
        "https://pt.wikipedia.org/wiki/Formiga-rainha"
    ],
    recomendacoes_de_amigos:[
        "https://pt.wikihow.com/Matar-uma-Formiga-Rainha"
    ],

    },
    ],

    }
}

```

Figura 12 – Quadro de Modelo de Dados
Fonte: Autor

6.2.2 LINGUAGEM DE PROGRAMAÇÃO

Para a linguagem de programação, foi escolhida o Python 3.8 para desenvolvimento da API, visto as bibliotecas disponíveis que auxiliam no tratamento de um grande número de dados e que auxiliam em diversas funcionalidades. Dentre as várias bibliotecas, as que foram essencialmente utilizadas neste projeto estão a

- *Scrape Search Engine*: possibilita a busca de termos em diversas Search Engines, como o Google, Bing e Yahoo, entre diversas outras, devolvendo as urls de resposta em formato JSON;
- *Urllib*: possibilita a abertura do conteúdo presente em sites, apenas passando a url como parâmetro;
- *BeautifulSoup*: permite a leitura de conteúdo presente em urls, realizando a remoção de *tags* encontradas em HTML5 e outras possíveis demais tecnologias;
- *NLTK*: possibilita a remoção de termos de um texto escrito em linguagem natural, fazendo o tratamento necessário como a remoção de *stopwords* e uso de *stemming*, transforma as palavras em tokens (um vetor de palavras) e realiza o cálculo do $TF \cdot IDF$;

Como o objetivo é criar uma aplicação Web, se fará uso do framework web python Flask (versão 1.1.2).

O Flask é um *Microframework*. Isso significa que ele não precisa de várias dependências para funcionar — ou seja, para colocar uma página na Web ou criar uma aplicação. Além disso, ele possui uma forma bem minimalista de utilizar o Python e é muito bem organizado.

Já para a construção da parte visual (Front-end), foi escolhido Javascript com a utilização do framework Vue.js. O Vue.js é um framework progressivo para a construção de interfaces de usuário. Sua biblioteca principal é focada exclusivamente na camada visual (*view layer*), tornando-o fácil de adotar e integrar com outras bibliotecas ou projetos existentes (VUE GUIDE, 2020).

6.4 ARQUITETURA DO SISTEMA

O modelo de dados do sistema prototípico pode ser representado pela seguinte imagem:

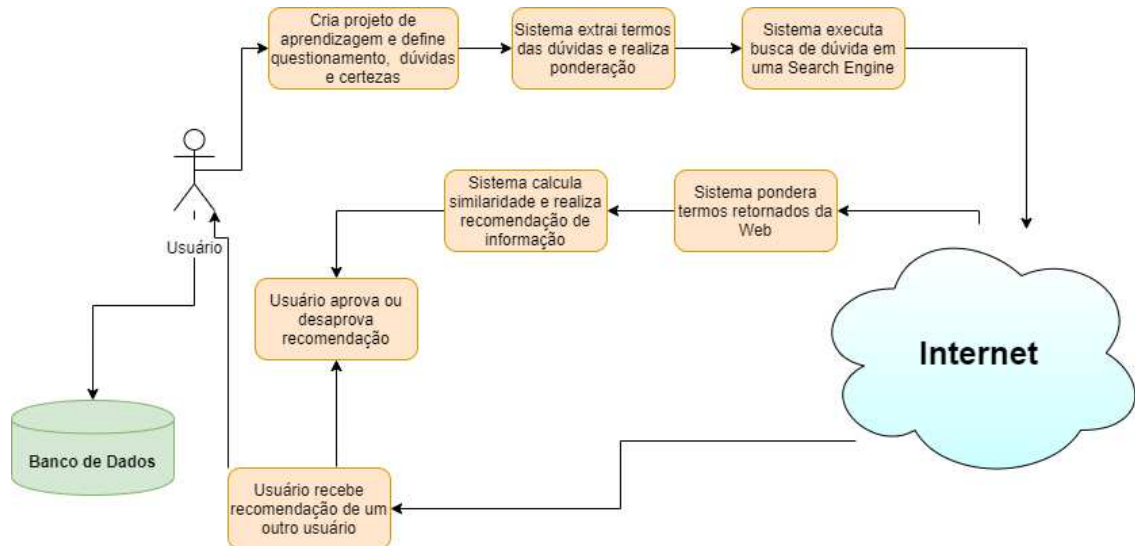


Figura 13 - Modelo de dados do sistema proposto

Fonte: Autor

Desse modo, ele deve seguir os seguintes comportamentos definidos:

1. O usuário deve criar um projeto de estudo (Projeto de Aprendizagem), onde deve definir o tema em estudo (questionamento), suas certezas e dúvidas;
2. As anotações realizadas de suas dúvidas deverão ser pré-processadas, entendidas e buscadas em documentos da Web;
3. O retorno dos documentos da Web também serão pré-processados e ponderados;
4. Será realizado o cálculo de similaridade entre os documentos retornados e as dúvidas do usuário e os cinco mais similares serão retornados como recomendação;
5. Ainda há a possibilidade da indicação direta, onde um amigo na rede social pode indicar documentos diretamente; *(passo opcional)*
6. Os documentos são sugeridos ao usuário e então podem ser aceitos ou não;
7. Os documentos aceitos servem como base para indicação direta a outro usuário.

6.5 RESULTADOS DOS TESTES

Para verificação da eficiência das atribuições do protótipo construído foram realizados testes para verificação das principais funcionalidades do sistema: recuperação da informação e recomendação da informação.

Para recuperação de informação, Salton e Buckley (1988, apud FARIAS, 2015) apresentam duas medidas para a avaliação da capacidade de um sistema para a recuperação de itens relevantes e rejeição dos itens com pouca importância. Essas medidas são a precisão e o *recall*.

Desse modo, para testes, foram simulados cinco projetos de aprendizagem com os seguintes questionamentos:

- Como funciona a organização de um formigueiro?
- Como funcionam os campos eletromagnéticos?
- Como é composto o sistema solar?
- Como acontece a metamorfose?
- Como ocorrem as correntes marítimas?

Foram definidas dúvidas e certezas para todos. O documento JSON com suas representações completas encontram-se no apêndice A e a descrição detalhada dos resultados dos testes encontram-se no apêndice B.

Desse modo, foi feita a ponderação dos termos da dúvida de cada uma das dúvidas do projeto de aprendizagem. Assim, a dúvida que possuiu a soma total de pesos maior, teve seus termos submetidos a busca. Os termos ponderados da dúvida selecionada também foram utilizados como parâmetro para o cálculo de similaridade com os documentos retornados. Para medir a relevância, foram verificados se os itens retornados respondiam ao questionamento principal, ou a alguma dúvida listada.

Apenas as primeiras cinco recomendações foram consideradas como itens recomendados.

Desse modo temos:

- Verdadeiros Positivos (VP): Retornados como recomendação e são relevantes;
- Falsos Negativos (FN): Não retornados como recomendação e são relevantes;
- Falsos Positivos (FP): Retornados como recomendação e são irrelevantes;

- Verdadeiros Negativos (VN): Não retornados como recomendação e são irrelevantes;

O cálculo da precisão compreende a proporção entre o número de itens relevantes recuperados e o número total de itens recuperados e pode ser obtida a partir da seguinte equação:

$$precisão = \frac{VP}{VP + FP}$$

Equação 4 - Cálculo da Precisão

O cálculo do recall considera a proporção de itens verdadeiros positivos em relação a todos esperados a serem considerados positivos. No entanto, não existe a possibilidade de considerar tal cálculo no ambiente proposto, visto que possuem em média em torno de dez itens retornados por busca e metade deles pode ser considerada como FN apenas pela limitação de retorno imposta, o que resultaria em um dado incorreto na simulação.

Sendo assim, foi realizado apenas o cálculo da precisão, obtendo os seguintes resultados:

Foram obtidos como resultado:

Questionamento	Número Total de Itens Recomendados	VP	FP	Precisão Obtida
Como funciona a organização de um formigueiro?	5	5	0	100%
Como funcionam os campos eletromagnéticos?	5	5	0	100%
Como é composto o sistema solar?	5	5	0	100%
Como acontece a metamorfose?	5	5	0	100%

Como ocorrem as correntes marítimas?	5	4	1	80%
--------------------------------------	---	---	---	-----

*Tabela 3 – Verificação da Precisão de Recomendações
Fonte: Autor*

Realizando a verificação da precisão no total itens retornados, independentes de serem recomendados, obtivemos:

Termo Pesquisado	Número Total de Itens Recuperados	VP	FP	Precisão Obtida
Como funciona a organização de um formigueiro?	11	9	2	81,81 %
Como funcionam os campos eletromagnéticos?	15	14	1	93,33%
Como é composto o sistema solar?	10	10	0	100%
Como acontece a metamorfose?	13	12	1	92,30%
Como ocorrem as correntes marítimas?	10	8	2	80%

*Tabela 4 – Verificação da Precisão de Itens Recuperados
Fonte: Autor*

O protótipo apresentado trabalha apenas com a remoção dos termos de maior frequência da área de interesse, dessa forma, embora os testes tenham apresentado bons resultados, por se tratar da mineração de textos de campos que possuem como entrada a digitação do usuário, os mesmos são sujeitos a erro, pois podem ter como entrada erros de digitação e siglas que podem ser removidas na verificação de *stopwords*.

7 CONSIDERAÇÕES FINAIS

Neste projeto foi abordada a perspectiva de se utilizar a tecnologia como ferramenta de auxílio ao autoestudo, de modo a prover acesso a uma gama de conhecimentos atuais sobre um determinado tópico escolhido, sendo capaz de assessorar o estudante na progressão de seu aprendizado ao longo da vida. Desse modo, objetivou-se a criação de uma ferramenta de ambiente pessoal de aprendizagem de forma análoga a um caderno inteligente para pessoas com perfil autônomo, de modo que estas fossem capazes de construir o seu próprio conhecimento.

Os ambientes pessoais de aprendizagem são capazes de fornecer ao aluno um papel ativo, colocando-o no centro do processo de aprendizado, ao mesmo tempo que aumenta o controle dos próprios alunos sobre o seu estudo. Assim, com essa assistência, eles podem gerenciar seu próprio ensino e trabalhar em direção a seus objetivos educacionais. A ferramenta proposta neste trabalho se propõe além dessa assistência, um auxílio durante o autoestudo do usuário por meio de recomendação de informações baseadas nos vestígios de informações deixadas pelo estudante durante seu autoestudo realizado por meio da pedagogia de Projetos de Aprendizagem.

Para construção dessa ferramenta, precisou-se analisar e escolher as melhores técnicas de recuperação e recomendação de informação para a necessidade proposta. Além do mais, foi necessário definir quais tecnologias atuais se adequassem melhor à construção do protótipo, bem como definir qual a melhor arquitetura e forma de comunicação entre seus componentes.

Após o estudo de possíveis tecnologias a serem implementadas no desenvolvimento do sistema proposto, foram realizados:

- A definição da especificação do modelo de dados utilizada;
- Definição da técnica de recuperação de informação;
- Definição da técnica de recomendação de informação;
- Desenvolvimento do protótipo para testes.
- Realização de Testes;

Após os testes, foi assimilada a capacidade do protótipo para realização de recuperação de informação e recomendação com base nas dúvidas do estudante. Desse modo é possível afirmar que o projeto cumpriu com os objetivos iniciais propostos e conseguiu cumprir

com a ideia inicial do projeto de desenvolver um protótipo para verificação seguindo as premissas objetivadas inicialmente.

7.1 TRABALHOS FUTUROS

Apesar dos resultados satisfatórios obtidos nesse trabalho, vale ressaltar que algumas melhorias podem ser desenvolvidas posteriormente. Algumas delas estarão descritas abaixo:

- Configuração para que ele possa realizar a recomendação da informação também por meio da filtragem colaborativa;
- Verificação de novas técnicas para recuperação da informação;
- Melhora na interface gráfica;
- Conexão com mais redes sociais.

REFERÊNCIAS

- AGGARWAL, C. C. **Recommender Systems: The Textbook**. (2016) Springer International Publishing Switzerland 2016. ISBN 978-3-319-29659-3 (eBook).
- ARIAS, J. O. C.; YERA, A. P. **O que é a Pedagogia Construtivista?**. Revista De Educação Pública, 5(8), 11-22. <https://doi.org/10.29286/rep.v5i8.360>. 2012.
- BARBOSA, C. E. M. **Estudo de Técnicas de Filtragem Híbrida em Sistemas de Recomendação de Produtos**. Monografia apresentada ao curso de Ciências da Computação da UFPE – Universidade Federal de Pernambuco. Recife. 2014.
- BERG, A. BUFFIE, E. F.; ZANNA, L-F. **IMF Working Paper - Should We Fear the Robot Revolution? (The Correct Answer is Yes)**. (2018) International Monetary Fund WP/18/116. Publicado em Maio de 2018.
- CASTRO, A. F. A.; BERTOLO, B. B.; BARVINSKI, C. A.; ODAKURA, V. **Ambiente Pessoal de Aprendizagem: Symbaloo.Edu**. (2015) Publicado em Computer on the Beach 2015 - Resumo Estendido.
- CAZELLA, S. C. ; DRUMM, J. ; BARBOSA, J. . **Um serviço para recomendação de artigos científicos baseado em filtragem de conteúdo aplicado a dispositivos móveis**. RENOTE. Revista Novas Tecnologias na Educação, v. 8, p. 22, 2010.
- COELHO, M. A. P.; DUTRA, L. R.; MARIELI, J. **ANDRAGOGIA e HEUTAGOGIA: práticas emergentes na educação**. Revista Transformar. Revista do Centro de Iniciação Científica e Extensão. Itaperuna, n.8, 2016.
- CRUZ, L. A. **Modelo para recuperação de informação em repositórios institucionais utilizando a técnica de sumarização a partir da seleção de atributos do Cassiopeia**. – Dissertação (Mestrado) – Programa de Pós-Graduação em Educação da Universidade Federal dos Vales do Jequitinhonha e Mucuri. Diamantina, MG, 2019.
- DE PAULA A. TRAMONTIN, Aline; GASPARINI, Isabela; PEREIRA, Roberto. **Sistemas de Recomendação com Elementos Sociais: Um Mapeamento Sistemático**. In: SIMPÓSIO BRASILEIRO DE SISTEMAS DE INFORMAÇÃO (SBSI), 14., 2018, Caxias do Sul. Anais

do XIV Simpósio Brasileiro de Sistemas de Informação. Porto Alegre: Sociedade Brasileira de Computação, nov. 2018. p. 30 - 23.

DINIZ, A. M. **Lifelong learning: aprendizado ao longo da vida.** (2017) Jornal Estadão. Disponível em: <https://educacao.estadao.com.br/blogs/ana-maria-diniz/lifelong-learning-aprendizado-ao-longo-da-vida/>. Publicado em 26 de Jan. de 2017. Acessado em 02 de Jun de 2019.

FAGUNDES, L. C.; NEVADO, R. A.; BASSO, M. V.; BITENCOURT, J. **Projetos de Aprendizagem – Uma Experiência Mediada por Ambientes Telemáticos.** Publicado em: Revista brasileira de informática na educação. Florianópolis. Vol. 14, no. 1 (jan./abr. 2006), p. 29-39, 2006.

FARIAS, M. P. **Um sistema de recomendação de links para formato de discussões de fóruns de um ambiente virtual de aprendizagem** – Dissertação (Mestrado) – Curso de Engenharia da Computação e Sistemas, Universidade Estadual do Maranhão – UEMA, 2015.

FERNEDA, E.; DIAS, G. A. **OntoSmart: um modelo de recuperação de informação baseado em ontologia.** (2017) Publicado em: Perspectivas em Ciência da Informação, v.22, n.2, p.170-187, abr./jun. 2017.

FLORES, A. M.; **A Informática na Educação: Uma Perspectiva Pedagógica** – Monografia – Universidade do Sul de Santa Catarina 1996. Disponível em: < <http://www.hipernet.ufsc.br/foruns/aprender/docs/monogr.htm> (nov/2002) >.

FONSECA, L.C.C, **Um Assistente Pessoal de Aprendizagem Continuada na WEB** – Tese Doutorado – Programa de Pós Graduação em Informática na Educação do Centro Interdisciplinar de Novas Tecnologias na Educação da Universidade Federal do Rio Grande do Sul – UFRS. Porto Alegre, 2009.

FONSECA, L. C. C.; FARIAS, M. P.; SILVA R. J. **Sistema de Recomendação de Links para o fomento de discussões em fóruns de um Ambiente Virtual de Aprendizagem.** RENOTE - Revista Novas Tecnologias na Educação. v. 12, n. 2 (2014).

HAWORTH, R. **Personal Learning Environments: A Solution for Self-Directed Learners.** (2016) Association for Educational Communications & Technology 2016. Publicado em 28 de Abr de 2016.

INTERNET SYSTEMS CONSORTIUM (2019). **Internet Domain Survey**. Disponível em: < <https://www.isc.org/survey/>>. Publicado em Jan de 2019. Acesso em Ago de 2020.

INTERNET WORLD STATS. **Internet Usage and Population Statistics for South America**. Publicado em 2020. Acesso em Ago de 2020.

JONASSEN, David. **O uso das novas tecnologias na Educação a Distância e a aprendizagem construtivista**. Em Aberto, Brasília, ano 16, n. 70, abr./jun. 1996, p. 70-88.

KOKAY, M. C. **Banco de dados NoSQL: Um novo paradigma** - Revista SQL Magazine 102 (2012). Disponível em: < <https://www.devmedia.com.br/banco-de-dados-nosql-um-novo-paradigma-revista-sql-magazine-102/25918> >. Publicado em 01 de Outubro de 2012. Acesso em Agosto de 2019.

KREBS, L. M.; ROCHA, R. P.; RIBEIRO, C. **Quem leu este também leu...: sistema de recomendação na biblioteca universitária**. (2017) Publicado em: Perspectivas em Ciência da Informação, v.22, n.1, p.151-169, jan./mar. 2017.

LIFELONG LEARNING COUNCIL QLD INC. **WHAT IS LIFELONG LEARNING?**. Disponível em: < <http://www.llcq.org/resources/>>. Acessado em 02 de Jun de 2019.

LONGWORTH, N.; DAVIES, W. K. **Lifelong Learning**. Edição: reimpressão. Routledge, 2014.

MONGODB. **CUSTOMER SUCCESS STORIES** [2008-2020]. Disponível em: < <https://www.mongodb.com/who-uses-mongodb>>. Acesso em Jan. de 2020.

MONTEIRO, S. D.; FERNANDES, R. P. M.; DECARLI, G. C.; TREVISAN, G. L. **Sistemas de recuperação da informação e o conceito de relevância nos mecanismos de busca: semântica e significação**. (2017) Publicado em: Encontros Bibli: revista eletrônica de biblioteconomia e ciência da informação, v. 22, n.50, p. 161-175, set./dez., 2017.

NETO, J. J.; FERNEDA, E. **ONTOLOGIA COMO RECURSO DE PADRONIZAÇÃO TERMINOLÓGICA NO PROCESSO DE RECUPERAÇÃO DE INFORMAÇÃO**. (2016) Publicado em: Inf. Pauta Fortaleza, CE v. 1 n. 1 jan./jun. 2016.

PINO, A. S.; RESENDE, **Teorias de aprendizagem na era digital: como e onde aprendemos?**. In: Emancipación digital: Políticas, prácticas educacionales e investigación. - 1a ed. - Montevideo: Universidad de la República; Flor de Ceibo, 2015. 216 p.

REZENDE, F. **As Novas Tecnologias na Prática Pedagógica Sob a Perspectiva Construtivista**. Publicado em Revista Ensaio, Belo Horizonte, v.02, n.01., p.70-87, jan-jun, 2002.

ŞAHİN, S.; ULUYOL, Ç. **Preservice Teachers' Perception and Use of Personal Learning Environments (PLEs)**. (2016) IRRODL - International Review of Research in Open and Distributed Learning. Volume 17, Number 2. Publicado em Fev de 2016.

SALADAGE, P. J.; FOWLER, M. **NoSQL Essencial: Um Guia Conciso para o Mundo Emergente da Persistência Poliglota**. (2019) Edição: Novatec Editora, 2019.

SANTOS, M. F. S. **A construção da autonomia do sujeito aprendiz no contexto da EaD**. ABED - Revista Brasileira de Aprendizagem Aberta e a Distância. v. 14. (2015).

SARDO, C.E. **Desempenho em Atividades de Auto-estudo por um Grupo de Alunos professores da Rede Pública do Estado Paraná**. Revista Expectativa. V.2, n.1 (2003).

SCHELMMER, E. **Projetos de aprendizagem gamificados: uma metodologia inventiva para a educação na cultura híbrida e multimodal**. Publicado em: Revista Momento Diálogos em Educação, v.27, n.1, p.42-69, 2018.

SILVA, Patrícia; DE MENEZES, Crediné; FAGUNDES, Lea. **Aprendizagem colaborativa: desenvolvimento de Projetos de Aprendizagem em ambientes digitais**. In: Anais do Workshop de Informática na Escola. 2016. p. 815.

SILVA, Reinaldo de Jesus da. **Sistema de recomendação de objeto de aprendizagem baseado em postagens extraídas do ambiente virtual de aprendizagem** – Tese (Doutorado) – Programa de Pós- Graduação em Informática na Educação do Centro Interdisciplinar de Novas Tecnologias na Educação, Universidade Federal do Rio Grande do Sul – UFRGS, 2016.

STROMQUIST, N. **World Development Report 2019: The changing nature of work**. World Bank. Washington, DC, World Bank, 2019, 151 pp. DOI:10.1596/978-1-4648-1328-3. Publicado por UNESCO Institute for Lifelong Learning and Springer Nature B.V. 2019.

VUE GUIDE. **O que é Vue.js?**. Disponível em: < <https://br.vuejs.org/v2/guide/> >. Acesso em Ago. de 2020.

APÊNDICE A – DADOS DE USUÁRIO USADOS PARA TESTES

```
{
  "usuario": {
    "_id": 1,
    "nome": "Leilanne",
    "projetos_de_aprendizagem": [
      {
        "questionamento": "Como funciona a organização de um
formigueiro?",
        "duvidas": [
          "Qual a quantidade de formigas em um formigueiro? ",
          "Como se escolhe a rainha de um formigueiro? ",
          "Existe apenas uma rainha no formigueiro? ",
          "Como as formigas se comunicam umas com as outras?
",
          "A rainha pode comandar as outras formigas, tomar
decisões? ",
          "Já foi comprovado algum nível de inteligência em um
formigueiro? ",
          "Que atividades inteligentes já foram identificadas
em formigueiros? "
        ],
        "certezas": [
          "Existe uma subdivisão de tarefas no formigueiro",
          "As formigas trabalham em função de proteger a
rainha",
          "As formigas responsáveis por buscar comida e manter
o formigueiro são as operárias",
          "As formigas podem se comunicar umas com as outras",
          "Existe uma hierarquia no formigueiro",
          "O formigueiro tem compartimentos com funções bem
definidas",
```

```

        "Um dos compartimentos do formigueiro é o depósito
de lixo"
    ],
    "recomendacoes_aceitas": [
        "https://www.infoescola.com/insetos/formigueiro/",
        "https://www.portaldosanimais.com.br/informacoes/tudo-sobre-
formigas-e-o-formigueiro/"
    ],
    "recomendacoes_recusadas": [
        "https://exame.com/ciencia/formigueiros-na-ilha-do-
cardoso-tem-ate-12-rainhas-por-colon/",
        "https://pt.wikipedia.org/wiki/Formiga-rainha"
    ],
    "recomendacoes_de_amigos": [
        "https://pt.wikihow.com/Matar-uma-Formiga-Rainha"
    ]
},
{
    "questionamento": "Como funcionam os campos
eletromagnéticos?",
    "duvidas": [
        "As linhas do campo eletromagnético são tangentes?",
        "Como as cargas se movem em um campo
eletromagnético?",
        "Como é a unidade de medida do campo
eletromagnético?"
    ],
    "certezas": [
        "É o ímã que cria o campo magnético",
        "O campo magnético terrestre protege a Terra da
radiação solar",

```

```

        "Campo eletromagnético é a concentração das cargas
        elétricas e magnéticas"
    ]
},
{
    "questionamento": "Como é composto o sistema solar?",
    "duvidas": [
        "Quais os planetas do sistema solar? ",
        "Quais outros astros encontramos no sistema solar?",
        "Como o sistema solar foi formado?"
    ],
    "certezas": [
        "É um conjunto de oito planetas",
        "Plutão é um planeta anão do sistema solar",
        "O sol é a estrela central do sistema solar",
        "A Terra é um pplaneta do sistema solar"
    ]
},
{
    "questionamento": "Como acontece a metamorfose?",
    "duvidas": [
        "A metamorfose pode ter mais de uma fase? ",
        "Quais os tipos de animais que sofrem metamorfose?",
        "Quanto tempo dura a metamorfose?"
    ],
    "certezas": [
        "A metamorfose é a transformação que acontece em
        alguns animais",
        "A borboleta sofre metamorfose",
        "A metarmofose modifica a estrutura do corpo do
        animal que a sofre"
    ]
},

```

```
{
  "questionamento": "Como ocorrem as correntes
marítimas?",
  "duvidas": [
    "O movimento de rotação da Terra tem influência em
relação às correntes marítimas? ",
    "As correntes marítimas seguem algum padrão?",
    "As correntes marítimas tem classificação?"
  ],
  "certezas": [
    "As correntes marítimas se originam na força do vento
na superfície da água",
    "Correntes marítimas são o movimento de grande massa
de água no oceano",
    "As correntes marítimas acontecem em todo mundo"
  ]
}
```

APÊNDICE B – DESCRIÇÃO CASOS DE TESTES

Questionamento	Como funciona a organização de um formigueiro?
Dúvidas	• Qual a quantidade de formigas em um formigueiro? • Como se escolhe a rainha de um formigueiro? • Existe apenas uma rainha no formigueiro? • Como as formigas se comunicam umas com as outras? • A rainha pode comandar as outras formigas, tomar decisões? • Já foi comprovado algum nível de inteligência em um formigueiro? • Que atividades inteligentes já foram identificadas em formigueiros?
Certezas	• Existe uma subdivisão de tarefas no formigueiro • As formigas trabalham em função de proteger a rainha • As formigas responsáveis por buscar comida e manter o formigueiro são as operárias • As formigas podem se comunicar umas com as outras • Existe uma hierarquia no formigueiro • O formigueiro tem compartimentos com funções bem definidas • Um dos compartimentos do formigueiro é o depósito de lixo
Dúvida submetida a engine search	Qual a quantidade de formigas em um formigueiro?
Recomendações Totais	1. http://revistagalileu.globo.com/Galileu/0,6993,ECT803829-3434,00.html 2. https://pt.wikipedia.org/wiki/Formiga

	3. https://www.linhaaid.com.br/pt-br/bug-id/ants/outdoor-mound-building-ants 4. https://super.abril.com.br/mundo-estranho/como-e-o-interior-de-um-formigueiro/ 5. https://super.abril.com.br/ciencia/a-dura-vida-das-formigas/ 6. https://super.abril.com.br/mundo-estranho/como-e-a-vida-dentro-de-um-formigueiro/ 7. https://pt.wikihow.com/Matar-uma-Formiga-Rainha 8. http://g1.globo.com/ciencia-e-saude/noticia/2015/10/em-colonia-de-formigas-25-dos-individuos-nao-trabalham-diz-estudo.html 9. https://educacao.uol.com.br/disciplinas/biologia/formigas-divisao-de-tarefas-e-cooperacao-fazem-parte-da-vida-de-insetos.htm 10. https://www.ddribeira.com.br/veneno-caseiro-para-formiga-19-remedios-caseiros-para-matar-formigas/ 11. https://www.bayerjovens.com.br/pt/materia/?materia=o-estranho-mundo-das-formigas
Recomendações consideradas irrelevantes	7 e 10

Questionamento	Como funcionam os campos eletromagnéticos?
Dúvidas	• As linhas do campo eletromagnético são tangentes? • Como as cargas se movem em um campo eletromagnético? • Como é a unidade de medida do campo eletromagnético?

Certezas	• É o ímã que cria o campo magnético • O campo magnético terrestre protege a Terra da radiação solar", • Campo eletromagnético é a concentração das cargas elétricas e magnéticas
Dúvida submetida a engine search	Como as cargas se movem em um campo eletromagnético?

Recomendações	1. https://www.educamaisbrasil.com.br/enem/fisica/forca-magnetica 2. https://www.educamaisbrasil.com.br/enem/fisica/eletromagnetismo 3. https://pt.wikipedia.org/wiki/Eletromagnetismo 4. https://pt.wikipedia.org/wiki/Campo_eletromagn%C3%A9tico 5. https://pt.wikipedia.org/wiki/Campo_magn%C3%A9tico 6. https://brasilecola.uol.com.br/fisica/eletromagnetismo.htm 7. https://www.todamateria.com.br/eletromagnetismo/ 8. https://mundoeducacao.uol.com.br/fisica/movimento-circular-no-campo-magnetico.htm 9. https://brasilecola.uol.com.br/fisica/cargas-eletricas-movimento.htm 10. https://mundoeducacao.uol.com.br/fisica/carga-movendo-no-campo-magnetico-uniforme.htm 11. https://brasilecola.uol.com.br/fisica/campo-magnetico.htm 12. https://www.alfaconnection.pro.br/fisica/eletromagnetismo/forca-magnetica/forca-sobre-carga-movel-em-campo-magnetico/ 13. https://www.todamateria.com.br/forca-magnetica/ 14. https://phet.colorado.edu/pt_BR/simulation/legacy/charges-and-fields 15. https://docplayer.com.br/7971298-Proposta-de-metodo-de-acesso-ao-meio-baseado-em-qos-para-redes-ad-hoc-ieee-802-11-tiago-trindade-da-silva.html
Recomendações consideradas irrelevantes	15

Questionamento	Como é composto o sistema solar?
----------------	----------------------------------

Dúvidas	• Quais os planetas do sistema solar? • Quais outros astros encontramos no sistema solar? • Como o sistema solar foi formado?
Certezas	• É um conjunto de oito planetas • Plutão é um planeta anão do sistema solar • O sol é a estrela central do sistema solar • A Terra é um planeta do sistema solar
Dúvida submetida a engine search	Quais os planetas do sistema solar?
Recomendações	1. https://www.todamateria.com.br/sistema-solar/ 2. https://www.stoodi.com.br/blog/fisica/sistema-solar-entenda/ 3. https://segredosdomundo.r7.com/planetas-do-sistema-solar/ 4. https://educacao.umcomo.com.br/artigo/quais-os-planetas-que-formam-o-sistema-solar-6047.html 5. https://www.hipercultura.com/planetas-sistema-solar/ 6. https://www.todamateria.com.br/planetas-do-sistema-solar/ 7. https://brasilecola.uol.com.br/geografia/planetas-2.htm 8. https://conhecimentocientifico.r7.com/sistema-solar-e-saiba-quais-planetas-o-compoem/ 9. https://mundoeducacao.uol.com.br/geografia/sistema-solar.htm 10. https://pt.wikipedia.org/wiki/Sistema_Solar
Recomendações consideradas irrelevantes	0
Questionamento	Como acontece a metamorfose?

Dúvidas	• A metamorfose pode ter mais de uma fase? • Quais os tipos de animais que sofrem metamorfose? • Quanto tempo dura a metamorfose?
Certezas	
Dúvida submetida a engine search	Quanto tempo dura a metamorfose?
Recomendações	1. https://brasilecola.uol.com.br/o-que-e/biologia/o-que-e-metamorfose.htm 2. https://meuartigo.brasilecola.uol.com.br/biologia/metamorfose.htm#:~:text=A%20transforma%C3%A7%C3%A3o%20ocorre%20em%20quatro,alguns%20dias%20at%C3%A9%20um%20m%C3%AAs. 3. https://mitologiaecultura.blogspot.com/2010/11/metamorfo.html 4. https://brasilecola.uol.com.br/biologia/metamorfose-nos-anfibios.htm 5. https://escolakids.uol.com.br/ciencias/metamorfose-das-borboletas.htm 6. https://mundoeducacao.uol.com.br/biologia/metamorfose-anfibios.htm 7. https://supernaturalbrasil.fandom.com/pt-br/wiki/Metamorfo 8. https://tsrd.fandom.com/pt-br/wiki/Metamorfo 9. http://shadow-falls.weebly.com/metamorfos.html 10. https://super.abril.com.br/ciencia/lagarta-leva-um-ano-para-virar-borboleta/ 11. https://site.veracruz.edu.br/blogs/6ano2019/2019/04/05/ciclo-de-vida-das-lagartas/ 12. https://br.answers.yahoo.com/question/index?qid=20070911051635AAw31yh

	13. https://www.portalsaofrancisco.com.br/curiosidades/metamorfose-da-borboleta
Recomendações consideradas irrelevantes	12

Questionamento	Como ocorrem as correntes marítimas?
Dúvidas	• O movimento de rotação da Terra tem influência em relação às correntes marítimas? • As correntes marítimas seguem algum padrão? • As correntes marítimas tem classificação?
Certezas	
Dúvida submetida a engine search	As correntes marítimas tem classificação?
Recomendações	1. https://www.portalsaofrancisco.com.br/geografia/correntes-maritimas 2. https://www.todoestudo.com.br/geografia/correntes-maritimas 3. https://brasilecola.uol.com.br/geografia/influencia-das-correntes-maritimas-no-clima.htm 4. http://www.geografia.seed.pr.gov.br/modules/galeria/detalhe.php?foto=1533&evento=5 5. https://exercicios.brasilecola.uol.com.br/exercicios-geografia/exercicios-sobre-influencia-das-correntes-maritimas-no-clima.htm 6. https://siteantigo.portaleducacao.com.br/conteudo/artigos/biologia/principais-correntes-marinhas/54462

	7. https://mundoeducacao.uol.com.br/geografia/correntes-maritimas.htm 8. https://brasilecola.uol.com.br/geografia/correntes-maritimas.htm 9. https://conteudo.tesouro.gov.br/manuais/index.php?option=com_content&view=article&id=1567:020332-classificacoes-orcamentarias&catid=749&Itemid=376 1. https://www.todamateria.com.br/correntes-maritimas/
Recomendações consideradas irrelevantes	5 e 9

APÊNDICE C – TELAS DO PROTÓTIPO DESENVOLVIDO

Menu

Recomendações

Como funciona a organização de um formigueiro?

Adicionar Anotação

Anotação

Tipo

+ Adicionar

Dúvidas

Anotações	Tags	Tornar certeza
Qual a quantidade de formigas em um formigueiro?		☐
Como se escolhe a rainha de um formigueiro?		☐
Existe apenas uma rainha no formigueiro?		☐
Como as formigas se comunicam umas com as outras?		☐
A rainha pode comandar as outras formigas, tomar decisões?		☐

Certezas

Anotações	Tags	Refutar
Existe uma subdivisão de tarefas no formigueiro		☐
As formigas trabalham em função de proteger a rainha		☐
As formigas responsáveis por buscar comida e manter o formigueiro são as operárias		☐
As formigas podem se comunicar umas com as outras		☐
Existe uma hierarquia no formigueiro		☐

Menu

Recomendações

Como funciona a organização de um formigueiro?

Adicionar Anotação

Anotação

Tipo

+ Adicionar

Dúvidas

Anotações	Tags	Tornar certeza
Qual a quantidade de formigas em um formigueiro?		☐
Como se escolhe a rainha de um formigueiro?		☐
Existe apenas uma rainha no formigueiro?		☐
Como as formigas se comunicam umas com as outras?		☐
A rainha pode comandar as outras formigas, tomar decisões?		☐

Certezas

Anotações	Tags	Refutar
Existe uma subdivisão de tarefas no formigueiro		☐
As formigas trabalham em função de proteger a rainha		☐
As formigas responsáveis por buscar comida e manter o formigueiro são as operárias		☐
As formigas podem se comunicar umas com as outras		☐
Existe uma hierarquia no formigueiro		☐

