

An Education Profile Model Applied in the Context of Recommender Systems

G. Bernardino, and A. Gonçalves

Abstract—The systems used in education have evolved bringing new teaching approaches and impacting in traditional learning methods. Also, they present new scenarios to students aiming to improve activities and practices in online systems. Along with this evolution the growing availability of contents has promoting an overload of data. Nowadays, many online educational systems have too much information, content and activities available to students making choices difficult. Educational systems supported by recommender systems allow to offer contents toward the needs and profiles of the students. This work presents a student profile model based on students' preferences and their interactions in online educational systems, with the purpose of allowing the recommendation of contents. To demonstrate the feasibility and evaluation of the proposed model two study scenarios were developed with information from an educational institution. In these scenarios recommendations were carried out using collaborative filtering and content based approaches. From the evaluation of these scenarios, the results demonstrate the viability of the model in producing recommendations that aim to aid the students' learning process when interacting with online educational systems.

Index Terms—Recommender Systems, Educational Systems, Student Profile.

I. INTRODUÇÃO

AS tecnologias da informação e comunicação cada vez mais desempenham um papel importante na popularidade dos sistemas online e vêm sendo aplicadas na educação de forma crescente [1]. Os sistemas educacionais evoluíram para grandes plataformas online com inúmeras atividades e mídias, utilizados por estudantes com diversos perfis e com objetivos diferentes [2].

A evolução dos sistemas educacionais também mudou o comportamento da aprendizagem tradicional apresentando novas situações aos estudantes e aprimorando as práticas nos sistemas online. Juntamente com esta evolução, o volume de dados armazenado nas bases é crescente impactando na sobrecarga de informações [3].

Neste volume de dados, porém, podem existir informações valiosas, gerando novos conhecimentos, possibilitando a identificação do perfil dos estudantes e sugerindo conteúdos e materiais que sejam adequados às suas atividades [3][4].

Este fato tem sido acompanhado por muitas pesquisas e estudos na área e cada vez mais surgem técnicas sendo aplicadas a estes sistemas, sendo muito significativo para a educação, pois diante de tantas inovações é necessário ter diferencial e buscar novas opções neste cenário [5].

Tendo em vista estes aspectos, surge o desafio em como despertar a atenção dos estudantes com conteúdos relevantes, associados ao seu perfil e às atividades educacionais. Desta forma, os sistemas de recomendação aplicados na educação podem trazer benefícios, trabalhando como filtros de informação, auxiliando os estudantes nos recursos educacionais e na seleção de conteúdos importantes associado com seu perfil [6][7].

Desde o início dos anos 2000 os Sistemas de Recomendação têm se tornado popular nos ambientes educacionais, visando auxiliar os estudantes na recomendação e personalização de cursos, atividades e materiais de aprendizagem [8]. Conforme a abordagem os Sistemas de Recomendação podem ser classificados em sistemas com filtragem baseada em conteúdo, baseada em conhecimento, filtragem colaborativa e híbrida, sendo estes os métodos mais tradicionais [9][10].

A eficiência dos Sistemas de Recomendação depende muito da precisão dos perfis do usuário, sendo este um item importante na proposta de um modelo de sistemas de recomendação, pois seus atributos são fundamentais para o processo [11][12]. Além disso, encontrar as abordagens mais promissoras é um desafio nestes sistemas, sendo que muitas pesquisas negligenciam o perfil do usuário [13].

Diante do exposto este trabalho tem como objetivo propor um modelo de perfil do estudante a partir de suas preferências e interações em sistemas educacionais *online*, com o propósito de permitir a recomendação de conteúdo. Para demonstração da viabilidade e avaliação do modelo proposto, foram desenvolvidos dois cenários de estudo com informações provenientes de uma instituição de ensino. Nestes cenários, foram aplicados processos de recomendações utilizando a abordagem de filtragem colaborativa e a abordagem baseada em conteúdo. Os resultados demonstram a viabilidade do modelo produzindo recomendações que visam auxiliar o processo de aprendizagem nos sistemas educacionais.

Este artigo está organizado em cinco seções, onde a seção 2 apresenta uma revisão da literatura dos principais conceitos deste trabalho, sendo Sistemas Educacionais, Sistemas de Recomendação, e os trabalhos correlatos. A seção 3 apresenta o modelo proposto e a seção 4 a aplicação e análise dos

G. S. Bernardino, Associação Beneficente da Indústria Carbonífera de Santa Catarina (SATC), Criciúma, Santa Catarina, Brasil (e-mail: giana.bernardino@satc.edu.br).

A. L. Gonçalves, Universidade Federal de Santa Catarina (UFSC), Araranguá, Santa Catarina, Brasil (e-mail: a.l.goncalves@ufsc.br).

resultados obtidos. Por fim, na seção 5 são apresentadas as considerações finais e trabalhos futuros.

II. REFERENCIAL TEÓRICO

A. Sistemas Educacionais

Sistemas educacionais são sistemas que proporcionam ao estudante um ambiente participativo, desperta o interesse e a curiosidade, estimula o raciocínio e a criatividade dos estudantes levando a uma aprendizagem significativa [14]. Os avanços das tecnologias da informação e comunicação e a explosão da internet na década de 90, trouxeram novas visões de ensino e de sistemas educacionais, resultando em mudanças radicais no modo de ensinar e aprender, apresentando novas possibilidades de ensino entre estudantes distantes geograficamente, transformando os sistemas educacionais em sistemas online [2].

O EaD (Ensino a Distância) surgiu com o avanço tecnológico ocasionando mudanças no processo de ensino-aprendizagem, transformando radicalmente os papéis dos educadores e estudantes, fazendo com que os sistemas educacionais se aprimorassem [15]. Como apoio a este processo a utilização dos Ambientes Virtuais de Aprendizagem (AVA) surgiu na educação, sendo atualmente utilizado em todas as modalidades de ensino [1].

Os AVAs ganharam espaço na educação tornando-se essenciais e utilizados por muitas instituições para ministrar, administrar e gerenciar cursos, disciplinas e aulas. Estes são sistemas computacionais que utilizam a internet como meio de acesso as atividades educacionais permitindo a integração de diversas mídias, linguagens e recursos e o armazenamento de todas as informações e interações em banco de dados [1].

Diante da diversificação de perfis de estudantes utilizando os AVAs, e estes sendo utilizados em todos os níveis de ensino, deve-se buscar uma personalização da instrução de acordo com as características do estudante, necessitando de sistemas educacionais inteligentes [16]. A construção do perfil é um aspecto muito importante nos ambientes educacionais, pois representa as informações individuais de cada estudante, incluindo características cognitivas, afetivas, sociais, perspectivas e feedback sobre os conteúdos. Desta forma, pode-se adaptar o sistema a suas necessidades [17].

Evidenciado em vários estudos o estilo de aprendizagem é um fator valioso no processo de ensino, como também na composição do perfil, pois indica como um estudante aprende e prefere aprender, mais especificadamente, como compreende a informação [18]. Os modelos de Felder-Silverman Learning Style Model (FSLSM) e de VARK, entre os modelos existentes, são empregados com êxito em diversos estudos e perfis educacionais, identificando as práticas e preferências de ensino de acordo com a necessidade dos estudantes [18]-[23].

O FSLSM [24] é um modelo de estilo de aprendizagem que pode ser capturado a partir do resultado de um questionário, podendo variar o estilo com o tempo, por isso, não devem ser considerados como características estáticas. O guia de estilos de aprendizagem VARK [25] é um questionário que auxilia os estudantes a determinarem sua modalidade preferida de

aprendizagem, sendo o objetivo sugerir estratégias que devem ser adotadas no aprendizado. Os resultados do modelo são indicativos e não diagnósticos.

B. Sistemas de Recomendação

Os Sistemas de Recomendação (SR) são ferramentas e técnicas que fornecem sugestões de itens de interesse ou relevância para um usuário [7][26]. Um dos principais objetivos dos SR é auxiliar os usuários nas tomadas de decisões, fornecendo recomendações de fácil acesso, alta qualidade e para um grande grupo de usuários, tornando está uma tecnologia poderosa e de impacto significativo [27].

Basicamente, os SR devem ser compostos por uma base de dados com as informações, dados de entrada representando os dados que o usuário produz ao interagir com o sistema e o algoritmo que combina as informações e os dados de entrada para prover sugestões [28].

Estes sistemas têm por objetivo reduzir a sobrecarga de informações e recuperar os objetos ou serviços mais relevantes, provendo deste modo recomendações personalizadas [29]. A personalização dos objetos é o diferencial dos sistemas de recomendação, pois permite respostas únicas e individualizadas aos usuários, através dos perfis, conteúdos, informações e feedback [30].

Uma das características mais importante deste tipo de sistema é a capacidade de prever preferências e interesses de um usuário através do seu comportamento e perfil. As preferências são fornecidas conforme os usuários interagem com os sistemas, sendo que nesta interação as informações são coletadas e utilizadas para construir o perfil [31].

De modo geral, existem duas formas de coletar as informações para a recomendação, de forma explícita ou implícita. Na forma explícita as informações podem ser coletadas através de questionários ou cadastros de informações que o usuário respondeu. Na forma implícita o sistema rastreia as atividades dos usuários e coleta as informações [27][31].

As abordagens mais tradicionais e aplicadas nos sistemas de recomendação são as com filtragem baseada em conteúdo, baseada em conhecimento, filtragem colaborativa e híbrida. Os sistemas com filtragem baseada em conteúdo sugerem itens que são semelhantes aos itens que o usuário preferiu no passado. Os sistemas com filtragem baseada em conhecimento sugerem itens para usuários identificando a correlação entre suas preferências e produtos, serviços ou conteúdo, em geral utilizando como suporte uma estrutura de representação do conhecimento. Os sistemas com filtragem colaborativa geram recomendações com base em opiniões de outros usuários que compartilharam interesses semelhantes. Já os sistemas híbridos combinam as melhores características de duas ou mais técnicas com o propósito de melhorar as recomendações [9][28].

Neste trabalho será utilizado no cenário de recomendação com abordagem colaborativa a técnica de medida de similaridade do cosseno por ter sido aplicada em diversas pesquisas e demonstrado resultados positivos e precisos [27][32][33]. A medida de similaridade do cosseno é uma métrica que mede a semelhança entre dois vetores n-

dimensionais com base no ângulo entre eles. O autor [27] define a semelhança entre dois itens a e b , vistos como vetores de classificação correspondente, conforme (1).

$$\text{sim}(\vec{a} * \vec{b}) = \frac{\vec{a} * \vec{b}}{|\vec{a}| * |\vec{b}|} \quad (1)$$

Os possíveis valores de similaridade estão entre 0 e 1, onde valores próximos de 1 indicam uma forte similaridade e valores próximos de 0 indicam uma fraca similaridade.

No cenário de recomendação baseada em conteúdo deste trabalho será utilizado o modelo de espaço vetorial e a normalização através da medida estatística *TF-IDF* (Term Frequency-Inverse Document Frequency). O *TF-IDF* compreende a relação entre a força interna de determinado termo (*TF*) pela sua capacidade de discriminação considerando a coleção de documentos (*IDF*) [27].

Para calcular o valor de frequência de termo normalizado $TF(i, j)$ divide-se o número absoluto de ocorrências do termo i no documento j , pelo número máximo de ocorrência dos termos no documento j .

$$TF(i, j) = \frac{\text{freq}(i, j)}{\text{maxfreq}(j)} \quad (2)$$

No cálculo da frequência inversa do documento para o termo i , tem-se o \log de N dividido por $n(i)$, sendo N o número de todos os documentos que constam na coleção e $n(i)$ o número de documentos em que o termo i aparece.

$$IDF(i) = \log \frac{N}{n(i)} \quad (3)$$

O peso final *TF-IDF* para um termo i em um documento j é calculado como o produto dessas duas medidas, $TF(i, j)$ e $IDF(i)$.

No modelo vetorial o documento é representado como um vetor dos pesos calculados para cada termo através de alguma medida, por exemplo, o *TF-IDF*, indicando que quanto mais elevado for o peso do termo maior é a sua significância para o documento.

C. Trabalhos Correlatos

Para análise e seleção de estudos relacionados com este trabalho e que apresentassem soluções para o modelo proposto foi realizada a pesquisa bibliográfica nas bases de dados da Scopus® e SBIE (Simpósio Brasileiro de Informática na Educação). Na base de dados da Scopus® foram selecionados os estudos do tipo artigo e documentos de conferência, com data de publicação maior que 2010, empregando o argumento de busca (recommender system*) e limitando a busca nas palavras-chave (e-learning, learning, adaptive e-learning, learning analytics, education*). Na base de dados SBIE foram selecionados os estudos empregando o argumento de busca (sistema* recomenda*).

Os trabalhos citados a seguir foram selecionados por estarem relacionados com sistemas de recomendação na área educacional e apresentarem informações dos perfis de usuário,

ilustrando os dados e atributos destes perfis. A Tabela I sintetiza os trabalhos apresentando as principais características e informações utilizadas no desenvolvimento destes estudos.

TABELA I
CARACTERÍSTICAS DOS PERFIS DE USUÁRIOS

Trabalhos	Características dos Perfis de Usuário
[3]	Perfil de Aprendizagem (idade, graduação, conhecimentos, objetivos de aprendizagem, aprendizagem exigidas, atividades).
[8]	O Perfil do Usuário contém atributos como área de estudo, experiência tecnológica, objetos de aprendizagem e plataforma de aprendizagem.
[10]	O perfil é composto por um conjunto de atividades que um estudante deve fazer, juntamente com o resultado obtido em cada atividade.
[11]	O perfil é criado a partir do conhecimento escolar, que contempla interesses de pesquisa, projetos de pesquisa, artigos publicados, notas e currículo.
[17]	Dados Pessoais (nome, data de nascimento, endereço, sexo, filiação, conhecimento prévio, usuário, senha); Dados de Desempenho (estilo cognitivo, aprendizagem, interação, habilidades adquiridas, habilidades de raciocínio e colaborativas); Histórico de Aprendizagem (avaliações, atividades e interações).
[18]	O perfil é composto por estilos de aprendizagem conforme modelo de FLSM, níveis de conhecimento e classificações de estudantes em grupo.
[19]	Dados Pessoais (nome, data de nascimento, sexo, raça, localização, visão, audição); Dados de Contexto (dispositivo de acesso, velocidade de acesso, navegador, sistema operacional); Estilos de aprendizagem do modelo FLSM; Também constam Dados de Personalidade, Dados Acadêmicos e Psicológicos.
[20]	O perfil contempla nível de experiência, conhecimento prévio, desempenho e estilo de aprendizagem conforme modelo FLSM.
[21]	O modelo se baseia no estilo de aprendizagem e conhecimento prévio dos estudantes. O estilo de aprendizagem é inspirado no modelo de VARK.
[22]	Perfil do Usuário (id, nome, último nome, data de nascimento, sexo, nível educacional, estilo aprendizagem); Objetos de Aprendizagem (id, título, link, estilo, tipo); O perfil combina os modelos de FLSM e VARK.
[23]	Informações básicas (nome, contato, filiação, sexo, idade, profissão, nível educacional, autenticação, acessibilidade); Além das informações básicas, o modelo é composto por conhecimento e desempenho do estudante, interesses, objetivos de aprendizagem, estilos cognitivos conforme modelo FLSM e análise de sentimentos.
[32]	O modelo é composto por três componentes: perfil do estudante, conhecimento do estudante e as preferências educacionais do estudante.

Durante a pesquisa pode-se observar que muitos trabalhos resumem estas informações de perfis sem valorizar os dados de entrada dos processos. Além disso, poucos trabalhos apresentam a estrutura de dados ou modelo lógico do sistema, sendo que a maioria enfoca na prática e aplicação das recomendações.

Dos trabalhos analisados, todos apresentam informações explícitas no perfil do usuário, como dados pessoais, preferências e interesses, sendo estas informações estáticas que dificilmente mudam no decorrer do tempo. Os dados dos cursos, atividades e objetos de aprendizagem são utilizados na

maioria dos perfis por estarem relacionados com o propósito principal destes estudos, recomendar conteúdos educacionais. As informações implícitas aparecem em poucos trabalhos, sendo que a presença deste tipo de informação torna o modelo dinâmico e o processo mais eficaz.

Neste estudo, o modelo proposto combina as principais informações dos perfis analisados, com informações explícitas e implícitas. O objetivo foi agregar os atributos mais utilizados nos estudos analisados, para obter um modelo dinâmico e abrangente, tornando este diferenciado por reunir todas estas informações em uma única estrutura de dados.

No modelo proposto também optou-se por utilizar os dois estilos de aprendizagem, o FLSM e o VARK, pois desta forma obtêm-se mais informações do usuário referente às suas preferências em aprender e compreender a informação, tornando o modelo mais robusto e auxiliando na similaridade dos usuários em relação ao estilo de aprendizagem e na recomendação de atividades.

Todas as informações utilizadas no modelo são de grande importância para a geração das recomendações, mas as informações implícitas possuem um diferencial, pois auxiliam na aplicação das abordagens de recomendação com resultados mais precisos e apropriados às preferências do usuário.

Na próxima seção é detalhado o modelo do perfil sendo apresentada a estrutura e o modelo lógico.

III. MODELO PROPOSTO

Para a construção do modelo proposto analisou-se os trabalhos correlatos na área, sendo observadas e comparadas as informações e características mais utilizadas e que pudessem contribuir na recomendação de conteúdo deste modelo. Na Fig. 1 é possível visualizar a Estrutura Analítica do Projeto (EAP) que representa a descrição do modelo proposto deste trabalho. O modelo é dividido em quatro módulos, sendo eles, identificação, aprendizagem, educacional e histórico.

O módulo Identificação possui as principais informações pessoais do usuário, caracterizados pelo código do usuário, nome, nome de contato, data de nascimento, e-mail, sexo, estado civil, escolaridade e idioma.

O módulo Aprendizagem tem como objetivo representar o estilo de aprendizagem do usuário em relação às suas preferências no processo de ensino, sendo caracterizado pelo estilo cognitivo, estilo de aprendizagem, acessibilidade e áreas de interesse.

No estilo cognitivo é contemplado as cinco modalidades de aprendizagem do modelo VARK, nas quais são, visuais, auditivas, leitura/escrita, cinestésicos e multimodais. No estilo de aprendizagem foi contemplado o modelo FLSM, sendo dividido em quatro dimensões, sensorial/intuitivo, visual/verbal, ativo/reflexivo e sequencial/global. A acessibilidade é classificada em auditiva, visual, auditiva/visual ou nenhuma. As áreas de interesse estão relacionadas com os interesses dos usuários em campos de estudo.

O módulo Educacional contempla os cursos, disciplinas e atividades, em que os usuários estão matriculados. As

atividades são compostas por conteúdos, apostilas, fórum, aulas virtuais, avaliações ou outras atividades relacionadas com o curso e disciplina que podem estar cadastradas no sistema.

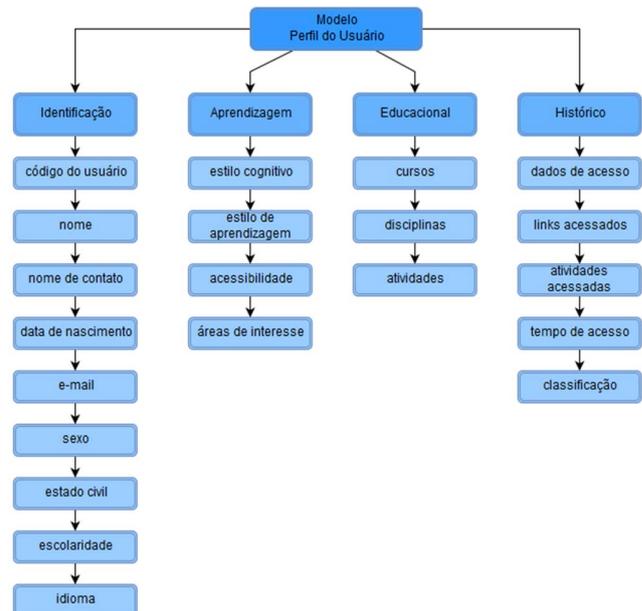


Fig. 1. Estrutura Analítica do Projeto (EAP).

O módulo Histórico representa as informações e interações que o usuário pode realizar no sistema educacional, sendo caracterizadas em dados de acesso, links acessados, atividades acessadas, tempo de acesso e classificação. A classificação é o feedback do usuário referente a atividade acessada. Os dados de acesso, neste modelo, contemplam o tipo de dispositivo que o usuário utiliza no acesso ao sistema, tipo de navegador (browser), sistema operacional e data. Nos links acessados é contemplado os menus e atividades que o usuário acessou no sistema. Nas atividades acessadas é prevista a descrição da atividade, hora de início e fim, resultando no tempo de acesso e a classificação que o usuário informou referente a atividade acessada.

As informações a serem utilizadas pelo modelo proposto podem ser coletadas de forma implícita ou explícita, dependendo da estrutura existente nos sistemas de gestão educacional dos cenários de aplicação.

O módulo Identificação pode selecionar todos os dados de um cadastro de usuário de forma implícita, caso exista no sistema de gestão, do contrário, deverá solicitar o preenchimento das informações de forma explícita. O módulo Aprendizagem deverá solicitar ao usuário que responda às suas preferências de aprendizagem tornando a captura das informações explícita. O módulo Educacional também poderá selecionar todas as informações de forma implícita quando existir no sistema de gestão educacional. Por fim, o módulo Histórico deverá capturar as informações de forma implícita e dinâmica, conforme o usuário interage com o sistema educacional *online*.

A partir da estrutura analítica do projeto foi gerado o

modelo de dados representado na Fig. 2 descrevendo os dados que pertencem ao domínio e que possam ser utilizados em um sistema de informação. O modelo lógico representa de forma abstrata a estrutura que o banco de dados possuirá.

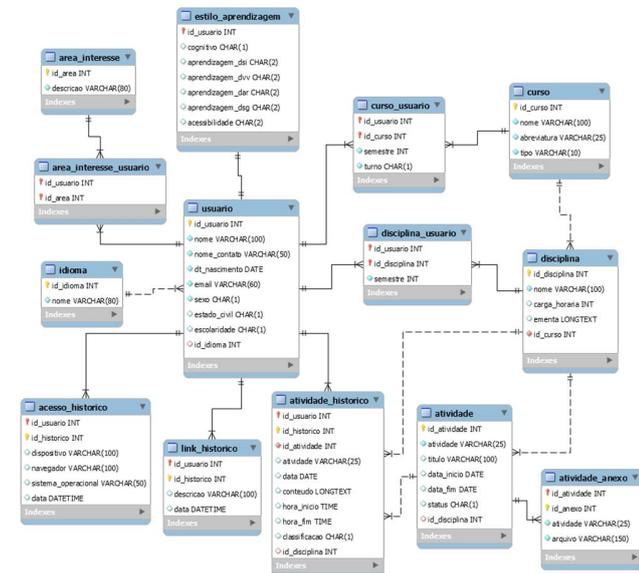


Fig. 2. Modelo Lógico.

Desta forma, construiu-se o modelo lógico conforme características citadas acima, formando o perfil do usuário para um sistema educacional. O modelo possui como propósito manter os dados coletados a partir das interações e acessos dos usuários em determinado ambiente virtual de aprendizado, com a finalidade de permitir a recomendação de conteúdos, bem como, de auxiliar o processo de aprendizagem.

IV. APLICAÇÕES E ANÁLISE DOS RESULTADOS

A. Cenário de Aplicação

Com o objetivo de avaliar o modelo proposto e demonstrar sua aplicabilidade e viabilidade, utilizou-se como cenário de aplicação um conjunto de dados da instituição de ensino SATC® (Associação Beneficente da Indústria Carbonífera de Santa Catarina) que dispõe dos seguintes níveis: educação infantil, fundamental, médio, técnico, técnico a distância, pós-técnico, graduação, especialização, mestrado e cursos livres.

Para a aplicação do cenário, foi criada uma base de dados conforme o modelo apresentado na Fig. 2. Para a importação optou-se em selecionar alguns cursos do técnico a distância e cursos livres por possuírem ambientes virtuais de aprendizagem com diversas atividades em execução. Foram importados 12 cursos considerando a modalidade técnico a distância e 47 cursos livres. Referente aos estudantes, foram importados 348 cadastros de estudantes e suas respectivas matrículas nos cursos, assim como, as disciplinas relacionadas.

Foram importadas as atividades que existiam cadastradas no ambiente virtual de aprendizagem, tais como, anotações virtuais, avaliações virtuais, conferências, conteúdos e fóruns,

totalizando 1004 atividades importadas. Em relação aos anexos das atividades, foram importados 1310 anexos.

Finalizando, foram importados os históricos de acesso dos estudantes referentes às atividades, links acessados e dados de acesso, conforme informações disponíveis no ambiente virtual de aprendizagem, sendo que algumas informações não existiam ficando nulas.

Desta forma, foram importadas e normalizadas todas as informações existentes nos sistemas da instituição de ensino utilizada para o cenário de aplicação. A seguir são detalhados os cenários e processos de recomendações.

B. Cenário de Recomendação com Abordagem Colaborativa

A proposta deste cenário é recomendar atividades aos estudantes matriculados no mesmo curso e disciplina com estilo de aprendizagem similar. Neste sentido, foi aplicada a abordagem colaborativa, pois tem como propósito recomendar itens a usuários semelhantes, utilizando a técnica de medida de similaridade do cosseno para medir a similaridade entre os usuários.

A primeira etapa deste processo consiste no cálculo dos pesos para cada atributo existente no estilo de aprendizagem do estudante para posteriormente, realizar o cálculo da similaridade. A partir das informações do estilo de aprendizagem, cada atributo recebe um peso produzindo dados auxiliares para a etapa de recomendação.

A segunda etapa consiste em calcular a similaridade dos estudantes conforme o cálculo da medida de similaridade do cosseno, detalhado na Seção 2. Foram selecionados 93 estudantes para o cálculo da similaridade que tinham as informações de estilo de aprendizagem. A similaridade foi calculada entre todos os estudantes selecionados, através de uma função gerando uma nova estrutura no banco de dados com a relação dos estudantes e o valor da similaridade entre eles, conforme Tabela II. Como resultado os valores de similaridade ficaram entre 0 e 1, onde valores próximos de 1 indicam uma forte similaridade e valores próximos de 0 indicam uma fraca similaridade.

TABELA II
RESULTADO DO CÁLCULO DA SIMILARIDADE

Curso	Disciplina	Estudante		Similaridade
		Origem	Destino	
3141	28489	46865	46702	0,973
3141	28489	46865	46741	0,958
3141	28489	46865	46809	0,912
4141	28497	32060	42071	0,99
4141	28497	32060	44053	0,95
4141	28497	32060	45290	0,815

A próxima etapa constituiu na seleção dos estudantes com similaridade maior ou igual a 0,9, bem como a análise das atividades que o estudante de origem não havia realizado, mas o estudante de destino sim. O limiar de 0,9 foi utilizado arbitrariamente com o intuito de demonstrar o cenário. A partir disso, foram produzidas as informações contendo a relação dos estudantes e as atividades não realizadas. Desta forma, as atividades a serem recomendadas aos estudantes com forte similaridade foram selecionadas, ou seja,

considerando determinado estudante de destino foram recomendadas atividades ainda não realizadas ou acessadas pelo estudante de origem. Na Tabela III observam-se as atividades recomendadas aos estudantes.

TABELA III
ATIVIDADES RECOMENDADAS AOS ESTUDANTES

Disciplina	Estudante de Origem	No. de Estudantes com Atividades Recomendadas	Título da Atividade Recomendada
28478	47911	22	Avaliação inicial
28488	46909	22	Unidade 1
28488	47005	21	Fórum Tira-Dúvidas
28488	40326	21	Trabalho de Mineralogia
28478	47911	21	Avaliação On-line
28478	47494	21	Aula 3 – MS Word

No total foram geradas 9276 recomendações, referente a 93 estudantes e 383 atividades relacionadas às disciplinas selecionadas neste cenário.

C. Cenário de Recomendação Baseada em Conteúdo

O objetivo deste cenário é a recomendação de cursos livres ofertados na instituição e relacionados às áreas de interesses dos estudantes. Neste cenário, foi aplicada a abordagem de recomendação baseada em conteúdo que tem como propósito explorar informações derivadas de texto e recomendar novos itens relacionados ao perfil do usuário.

Para a normalização dos dados foi utilizado o cálculo do *TF-IDF*, conforme demonstrado na Seção II. Para a recomendação dos cursos livres foram analisadas as ementas das disciplinas de cada curso em relação às áreas de interesse.

A primeira etapa deste processo constituiu na criação de uma nova estrutura de dados com os termos pertencentes às áreas de interesse cadastradas no modelo, excluindo-se as preposições. Após isso, a próxima etapa calculou o *TF-IDF* de cada termo em relação a cada ementa de disciplina do curso. Através de funções foram calculados os atributos e armazenados em uma nova estrutura de dados com o peso de cada termo, conforme Tabela IV.

TABELA IV
ATRIBUTOS E NORMALIZAÇÃO DO TF-IDF

FQ	MFQ	NDIS	NDIST	TF	IDF	TF-IDF
1	1	69	1	1,00	4,23	4,23
1	2	69	5	0,50	2,62	1,31
6	6	69	5	1,00	2,62	2,62
2	4	69	5	0,50	2,62	1,31
1	1	69	5	1,00	2,62	2,62
1	9	69	10	0,11	1,93	0,21
2	2	69	10	1,00	1,93	1,93

Para o cálculo do TF quantificou-se o número de vezes que um termo aparece em uma ementa de uma disciplina, sendo denominado neste cenário de atributo FQ. Também foi quantificada a frequência do termo mais frequente na respectiva disciplina, sendo este, o termo de maior frequência na ementa da disciplina, denominado atributo MFQ. O valor de TF é calculado através dos atributos FQ dividido por MFQ.

Para o cálculo do IDF foi quantificado o total de disciplinas recomendáveis, em relação aos cursos livres selecionados, denominado atributo NDIS. Também foi quantificado o número de disciplinas em que o termo aparece, denominado NDIST. A partir dessas variáveis calculou-se o atributo TF-IDF, sendo este a relevância de cada termo em relação à ementa da disciplina dos cursos, sendo que os termos com pesos maiores possuem maior importância na recomendação em relação aos pesos menores.

A última etapa deste processo teve o objetivo de selecionar os estudantes que tinham registrado em seu perfil as áreas de interesse de sua preferência e que estavam relacionadas com as disciplinas e cursos a serem recomendados. Nesta etapa, também foi gerada uma nova estrutura com as recomendações aos estudantes, os cursos livres que estavam relacionados com as suas áreas de interesse e com o peso do *TF-IDF* maior que zero.

Abaixo, na Tabela V, pode-se observar os cursos livres que foram recomendados aos estudantes conforme áreas de interesse de sua preferência. Foram selecionadas as recomendações ordenando-as do maior peso (*TF-IDF*) para o menor peso, desta forma, as primeiras linhas representam as recomendações mais significativas do processo.

TABELA V
CURSOS LIVRES RECOMENDADOS AOS ESTUDANTES

Estudante	Área	Termo	TF-IDF	Descrição do Curso
36873	Processamento Digital de Sinais	Digital	4,23	Photoshop
45747	Processamento Digital de Sinais	Digital	4,23	Photoshop
24621	Eletrônica de Potência e Acionamento	Eletrônica	4,23	PPCP – Planejamento
36873	Eletrônica de Potência e Acionamento	Eletrônica	4,23	PPCP – Planejamento
46416	Decoração e Paisagismo	Decoração	3,54	Decoração de Interiores
46414	Normas (ABNT)	ABNT	3,54	Eletricidade Predial
46414	Normas (ABNT)	ABNT	3,54	NR10 Segurança em Instalações

Destaca-se que neste cenário as recomendações foram geradas a partir do peso de cada termo individualmente, neste caso, pode-se observar as duas primeiras linhas da Tabela V, onde o termo Digital teve um peso significativo, recomendando o curso de Photoshop aos estudantes que tinham selecionado a área de interesse Processamento Digital de Sinais em seu registro. Desta forma, o termo Digital analisado individualmente, possui relação com o curso de Photoshop. Contudo, caso seja analisada a área de interesse Processamento Digital de Sinais com todos os termos, esta não possui uma relação significativa com o curso recomendado, sendo mais efetivo para este tipo de cenário calcular e analisar a área de interesse com todos os termos.

Após o cálculo, obteve-se como maior peso (*TF-IDF*) o valor de 4,23 e o menor peso o valor de 0,21. Foram geradas um total de 761 recomendações referente a 30 estudantes, 19 áreas de interesse e 38 cursos.

V. CONSIDERAÇÕES FINAIS

Neste trabalho foi apresentado um modelo de perfil do estudante, que a partir de suas características, preferências e interações teve o propósito de permitir a recomendação de conteúdos em sistemas educacionais online. A fim de atingir o objetivo foi realizada uma revisão da literatura e analisados os trabalhos relacionados às áreas de Sistemas de Recomendação e Sistemas Educacionais. A partir desta etapa, foram selecionados e delimitados os dados utilizados na obtenção do perfil do estudante, assim como, as abordagens utilizadas na recomendação dos conteúdos.

Durante a pesquisa pode-se observar que, dos trabalhos correlatos, poucos apresentam a estrutura ou modelo lógico do sistema. A maioria enfoca na prática e aplicação das recomendações, abordagens e técnicas sem detalhar o modelo de dados utilizado para suportar a geração das recomendações.

Para a construção do modelo proposto foi desenvolvida a descrição do modelo representado através de uma EAP, bem como, o modelo lógico. Com o objetivo de demonstrar a viabilidade e analisar o modelo proposto, foi definido um banco de dados e foram importados os dados. Para efetividade do modelo foram aplicados dois cenários de recomendação utilizando a abordagem de filtragem colaborativa e versão simplificada da abordagem baseada em conteúdo. Nos cenários aplicados foi possível gerar as recomendações, sendo que no primeiro cenário foram recomendadas atividades aos estudantes conforme similaridade e acesso. No segundo cenário foram recomendados cursos livres conforme área de interesse dos estudantes.

Analisando os resultados e processo de recomendação dos cenários, pode-se verificar que o modelo proposto foi capaz de atender ao objetivo desta pesquisa em recomendar atividades e cursos aos estudantes conforme interações e preferências. Os cenários propostos e aplicados obtiveram os resultados que podem ser considerados apropriados em relação aos conteúdos recomendados, ou seja, produziram recomendações que demonstram serem capazes de auxiliar os estudantes nos sistemas educacionais online no que tange às atividades e aos cursos.

Como perspectivas de trabalhos futuros vislumbra-se a utilização no modelo proposto de outras abordagens de recomendação que não foram aplicadas nesta pesquisa, como as abordagens baseada em conhecimento, demográfica e a abordagem híbrida, que seria a combinação de duas ou mais abordagens.

Referente aos cenários de recomendação, o modelo proposto permite aplicar outros processos de recomendação, por exemplo, recomendar as atividades relacionadas ao estilo de aprendizagem e anexos das atividades, recomendar as atividades relacionadas à acessibilidade do usuário, entre outros.

REFERÊNCIAS

- [1] R. E. D. Silva, J. L. C. Ramos, R. L. Rodrigues, A. S. Gomes, and A. V. Fonsêca, "Mineração de dados educacionais na análise das interações dos alunos em um Ambiente Virtual de Aprendizagem," *Simpósio Brasileiro de Informática na Educação*, no. Sbie, pp. 1197–1206, 2015.
- [2] V. G. Machado, "Aula Virtual: implicações e desafios docentes considerando o cenário de educação apoiada por tecnologias digitais," Pontifícia Universidade Católica do Rio Grande do Sul, 2013.
- [3] D. Wu, J. Lu, and G. Zhang, "A Fuzzy Tree Matching-Based Personalized E-Learning Recommender System," *IEEE Transactions on Fuzzy Systems*, vol. 23, no. 6, pp. 2412–2426, 2015.
- [4] J. L. C. Ramos, R. E. D. Silva, R. L. Rodrigues, J. C. S. Silva and A. S. Gomes. A Comparative Study between Clustering Methods in Educational Data Mining, *IEEE Latin America Transactions*, vol. 14, no. 8, pp. 3755-3761, 2016.
- [5] M. H. Ansari, M. Moradi, O. Nikrah, and K. M. Kambakhsh, "CodERS: A hybrid recommender system for an E-learning system," *Proc. - 2016 2nd Int. Conf. Signal Process. Intell. Syst. ICSPIS 2016*, pp. 14–15, 2017.
- [6] E. Durán and M. Alvarez. Recommendations of Experts to advise students on ubiquitous learning experiences: Method and Validation, *IEEE Latin America Transactions*, vol. 16, no. 9, pp. 2314-2320, 2018.
- [7] A. Gordillo, E. Barra and J. Quemada, "A Hybrid Recommendation Model for Learning Object Repositories," *IEEE Latin America Transactions*, vol. 15, no. 3, pp. 462-473, 2017.
- [8] A. Zapata, V. M. Menendez, M. E. Prieto, and C. Romero, "A hybrid recommender method for learning objects," *IJCA Proc.*, pp. 1–7, 2011.
- [9] W. Carrer-Neto, M. L. Hernández-Alcaraz, R. Valencia-García, and F. García-Sánchez, "Social knowledge-based recommender system. Application to the movies domain," *Expert Syst. Appl.*, vol. 39, no. 12, pp. 10990–11000, 2012.
- [10] J. Serrano-Guerrero, F. P. Romero, and J. A. Olivas, "Hiperion: A fuzzy approach for recommending educational activities based on the acquisition of competences," *Inf. Sci. (Nj)*, vol. 248, pp. 114–129, 2013.
- [11] B. Amini, R. Ibrahim, M. S. Othman, and A. Selamat, "Capturing scholar's knowledge from heterogeneous resources for profiling in recommender systems," *Expert Syst. Appl.*, vol. 41, no. 17, pp. 7945–7957, 2014.
- [12] N. D. D. Méndez, "Modelo Adaptativo Multi-Agente para la Planificación y Ejecución de Cursos Virtuales Personalizados," Cambridge University Press, Cambridge, 2009.
- [13] J. Beel, B. Gipp, S. Langer, and C. Breiteringer, "Research-paper recommender systems: a literature survey," *Int. J. Digit. Libr.*, vol. 17, no. 4, pp. 305–338, 2015.
- [14] J. A. T. Almeida and P. R. Ross, "Softwares Educativos: Alternativas para Aprender e Ensinar em Sala de Recursos Multifuncional - Tipo I," *Os Desafios da Escola Pública Paranaense na Perspectiva do Professor PDE - Cadernos PDE*, vol. 1, 2014.
- [15] F. L. Matuchski and P. T. Lupion, "Potencialidades e limitações do ambiente virtual de aprendizagem em um curso online," *Revista Intersaberes*, vol. 5, no. 10, pp. 152–166, 2010.
- [16] I. I. Bittencourt, P. Brito, A. Pedro, S. Isotani, P. A. Jaques, and C. Rubira, "Desafios da Engenharia de Software na Educação: Variabilidade de Sistemas Educacionais Inteligentes e Instanciação em Larga Escala," in *Workshop de Desafios da Computação Aplicada à Educação*, 2012, pp. 1–10.
- [17] B. Vesin, A. Klačnja-Milićević, M. Ivanović, and Z. Budimac, "Applying recommender systems and adaptive hypermedia for e-learning personalization," *Comput. Informatics*, vol. 32, no. 3, pp. 629–659, 2013.
- [18] P. Dwivedi and K. K. Bharadwaj, "e-Learning recommender system for a group of learners based on the unified learner profile approach," vol. 32, no. 2, pp. 293–312, 2015.
- [19] H. M. G. Gutierrez, N. D. D. Mendez, and D. A. O. Carranza, "Modelo del Estudiante para Sistemas Adaptativos de Educación Virtual," *Rev. Av. en Sist. e Informática*, vol. 5, no. 1, pp. 199–206, 2008.
- [20] H. Imran, M. Belghis-Zadeh, T.-W. Chang, Kinshuk, and S. Graf, "PLORS: a personalized learning object recommender system," *Vietnam J. Comput. Sci.*, vol. 3, pp. 3–13, 2016.
- [21] A. Mavroudi and T. Hadzilacos, "Broadening the use of e-learning standards for adaptive learning," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 7558 LNCS, pp. 215–221, 2012.
- [22] P. Rodríguez, V. Tabares, N. Duque, D. Ovalle, and R. M. Vicari, "Multi-agent Model for Searching, Recovering, Recommendation and Evaluation of Learning Objects from Repository Federations," *Multagent Syst.*

- based Learn. Environ.*, pp. 631–640, 2012.
- [23] K. Verbert *et al.*, “Context-aware recommender systems for learning: A survey and future challenges,” *J. Latex Cl. Files*, vol. 6, no. 1, pp. 2–3, 2007.
- [24] R. Felder and L. Silverman, “Learning and teaching styles in engineering education,” *Eng. Educ.*, vol. 78, no. June, pp. 674–681, 1988.
- [25] N. Fleming, E. Developer, D. Baume, and H. E. Consultant, “Learning Styles Again : VARKing up the right tree !,” *Educ. Dev.*, no. 7, pp. 4–7, 2006.
- [26] C. Figueroa, I. Vagliano, O. R. Rocha, and M. Morisio, “A systematic literature review of Linked Data-based recommender system,” *Concurr. Comput. Pract. Exp.*, vol. 27, no. 6, pp. 4659–4684, 2015.
- [27] D. Jannach, M. Zanker, A. Felfernig, G. Friedrich, *Recommender Systems An Introduction*. 2011.
- [28] R. Burke, “Hybrid Recommender Systems: Survey and Experiments,” *User Model. UserAdapted Interact.*, vol. 12, no. 4, pp. 331–370, 2002.
- [29] J. Lu, D. Wu, M. Mao, W. Wang, and G. Zhang, “Recommender system application developments: A survey,” *Decis. Support Syst.*, vol. 74, pp. 12–32, 2015.
- [30] R. Noor and F. A. Khan, “Personalized recommendation strategies in mobile educational systems,” *2016 6th Int. Conf. Innov. Comput. Technol. INTECH 2016*, pp. 435–440, 2016.
- [31] J. Bobadilla, F. Ortega, A. Hernando, and A. Gutiérrez, “Recommender systems survey,” *Knowledge-Based Syst.*, vol. 46, pp. 109–132, 2013.
- [32] M. K. Khribi, M. Jemni, and O. Nasraoui, “Automatic Recommendations for E-Learning Personalization Based on Web Usage Mining Techniques and Information Retrieval,” *J. Educ. Technol. Soc.*, vol. 12, no. 4, pp. 30–42, 2009.
- [33] F. Ricci, L. Rokachm, B. Shapira, and P. B. Kantor, *Recommender System Handbook*, vol. 532. 2009.



Giana da Silva Bernardino possui graduação em Ciência da Computação pela Universidade do Sul de Santa Catarina (2000), especialização em MBA Gerenciamento de Banco de Dados pela Universidade do Extremo Sul Catarinense (2005) e mestrado em Tecnologias da Informação e Comunicação pela UFSC (2017). Atualmente é Analista de

Sistemas da Associação Beneficente da Indústria Carbonífera de Santa Catarina - SATC e Professora de Ensino Superior do curso de Tecnologia em Análise e Desenvolvimento de Sistemas, modalidade EAD semi presencial. Tem experiência na área de Análise e Desenvolvimento de Sistemas e Banco de Dados. <http://lattes.cnpq.br/0464576625190425>.



Alexandre Leopoldo Gonçalves possui graduação em Ciência da Computação pela Fundação Universidade Regional de Blumenau (1997), mestrado e doutorado em Engenharia de Produção pela Universidade Federal de Santa Catarina em 2000 e 2006. Atualmente é Professor Adjunto lotado no Campus Araranguá/UFSC e Professor Permanente

do Programa de Pós-Graduação em Engenharia e Gestão do Conhecimento/UFSC. Tem experiência nas áreas de Ciência da Computação e Engenharia do Conhecimento atuando principalmente nos seguintes temas: Extração e Recuperação de Informação, Descoberta de Conhecimento e Engenharia de Ontologia. <http://lattes.cnpq.br/5138758521691630>.