



**UNIVERSIDADE FEDERAL DE SANTA CATARINA
CENTRO DE CIÊNCIAS, TECNOLOGIAS E SAÚDE
PROGRAMA DE PÓS-GRADUAÇÃO EM
TECNOLOGIAS DA INFORMAÇÃO E COMUNICAÇÃO**

Mayara Sousa Stein

**ANÁLISE QUANTITATIVA DA CONCORDÂNCIA
ENTRE USUÁRIOS AVALIADORES DE OBJETOS DE
APRENDIZAGEM: UM ESTUDO DE CASO COM
DADOS DO REPOSITÓRIO MERLOT**

Araranguá

2020

Mayara Sousa Stein

**ANÁLISE QUANTITATIVA DA CONCORDÂNCIA
ENTRE USUÁRIOS AVALIADORES DE OBJETOS DE
APRENDIZAGEM: UM ESTUDO DE CASO COM
DADOS DO REPOSITÓRIO MERLOT**

Dissertação submetida ao Programa de Pós-Graduação em Tecnologias da Informação e Comunicação para a obtenção do Grau de Mestre em Tecnologias da Informação e Comunicação.
Orientador: Prof. Cristian Cechinel, Dr.
Coorientador: Prof. Vinicius F. C. Ramos, Dr.

Araranguá

2020

Ficha de identificação da obra elaborada pelo autor,
através do Programa de Geração Automática da Biblioteca Universitária da UFSC.

Stein, Mayara Sousa

Análise quantitativa da concordância entre usuários avaliadores de objetos de aprendizagem: um estudo de caso com dados do repositório Merlot / Mayara Sousa Stein ; orientador, Cristian Cechinel, coorientador, Vinicius Ramos, 2021.

104 p.

Dissertação (mestrado) - Universidade Federal de Santa Catarina, Campus Araranguá, Programa de Pós-Graduação em Tecnologias da Informação e Comunicação, Araranguá, 2021.

Inclui referências.

1. Tecnologias da Informação e Comunicação. 2. Qualidade. 3. Objetos de Aprendizagem. 4. Repositório de objetos e aprendizagem. 5. Concordância. I. Cechinel, Cristian. II. Ramos, Vinicius. III. Universidade Federal de Santa Catarina. Programa de Pós-Graduação em Tecnologias da Informação e Comunicação. IV. Título.

**Análise Quantitativa da Concordância entre Usuários
Avaliadores de Objetos de Aprendizagem: Um Estudo de
Caso com Dados do Repositório Merlot**

O presente trabalho em nível de mestrado foi avaliado e aprovado por banca examinadora composta pelos seguintes membros:

Prof.(a) Patricia Jantsch Fiuza, Dra.
Universidade Federal de Santa Catarina

Prof. Robson Rodrigues Lemos, Dr.
Universidade Federal de Santa Catarina

Prof.(a) Juliana Cristina Braga, Dra.
Universidade Federal do ABC

Certificamos que esta é a **versão original e final** do trabalho de conclusão que foi julgado adequado para obtenção do título de mestre em Tecnologias da Informação e Comunicação

Prof. Fernando José Spanhol, Dr.
Coordenação do Programa de Pós-Graduação

Prof. Cristian Cechinel, Dr.
Orientador

Prof. Vinicius F. C. Ramos, Dr.
Coorientador

Araranguá, 13 de novembro de 2020.

Dedico este trabalho a todos que estiveram presente na minha vida durante esta caminhada.

AGRADECIMENTOS

Agradeço, primeiramente, a Deus, pelas oportunidades e conquistas que eu consegui alcançar.

Agradeço aos meus familiares pelo apoio, em especial aos meus queridos pais Paulo Sérgio Stein e Rosária Ferreira Sousa, que sempre foram muito presentes na minha vida e nunca deixaram de apoiar as minhas decisões. A minha irmã Mayza Sousa Stein que esteve presente em todos os momentos durante esses dois anos e ao meu irmão Maicon Alexandre Sousa Stein.

Ao meu namorado Matheus Frascisco Batista Machado que sempre me apoiou durante toda essa caminhada e não mediu esforços quando mais precisei.

Meu imenso agradecimento ao meu orientador Cristian Cechinel, e ao meu coorientador, professor Dr. Vinicius F. C. Ramos pela dedicação e apoio durante a execução deste trabalho. Me orientaram com paciência e precisão, meus sinceros agradecimentos e admiração.

Ao Programa de Pós-Graduação em Tecnologias da Informação e Comunicação da Universidade Federal de Santa Catarina (Campus Araranguá) pela oportunidade e a todos os professores do PPGTIC pelos ensinamentos e colaborações.

Agradeço a banca examinadora pela disponibilidade e pelas significativas contribuições a esta pesquisa.

Enfim, aos amigos e colegas, pelo incentivo e apoio.

RESUMO

Com o aumento contínuo do uso de objetos de aprendizagem é importante que os repositórios de objetos de aprendizagem (ROA) adotem estratégias para a garantia da qualidade dos recursos. Exemplos dessas estratégias são o sistema de avaliação por pares de especialistas, avaliação realizada pelos usuários com pontuação e/ou comentários, premiação do editor, entre outros. Por exemplo, o Merlot busca garantir a qualidade dos recursos por meio de avaliações e pontuações fornecidas por pares-revisores (especialistas) e pela comunidade de usuários de diferentes áreas. Como podem ocorrer discrepâncias na avaliação da qualidade dos objetos de aprendizagem, é necessário entender o grau de concordância entre os usuários que são avaliadores no Merlot. Diante disso, o presente trabalho tem como objetivo analisar a concordância e a existência de associações entre os usuários que realizaram avaliações dentro do repositório Merlot. Para isso, foram realizados dois estudos de caso com dados coletados no Merlot. O primeiro estudo de caso avaliou a concordância entre pares de avaliadores que classificaram o mesmo conjunto de recursos considerando suas disciplinas preferidas e área de especialização. O segundo estudo de caso mediu a associação entre as classificações das médias das notas fornecidas pelos avaliadores provenientes de diferentes categorias de disciplina. Os resultados apontaram que os usuários avaliadores apresentam grande divergência nas avaliações fornecidas para o mesmo recurso. Além disso, as classificações das médias dos recursos considerando os avaliadores de diferentes categorias de disciplina não mostraram correlação. Os resultados aqui encontrados podem ser utilizados para aprimorar os mecanismos de garantia da qualidade dentro dos repositórios, bem como para o desenvolvimento de novas formas de apresentação das informações relacionadas à qualidade dos recursos.

Palavras-chave: Qualidade. Repositório de objetos e aprendizagem. Objetos de Aprendizagem. Concordância.

ABSTRACT

With the continuous increase in learning objects usage it is important that learning object repositories (LOR) adopt measures to assure the quality of their resources. Examples of these strategies are the peer review system, the assessment carried out by users with scores or/and comments, editor awards, among others. For instance, Merlot seeks to ensure the quality of the resources by using evaluations and scores provided by peer-reviewers (experts) and the community of users from different areas. As discrepancies in the assessment of the quality of learning objects may occur, it is necessary to understand the degree of agreement between users that are evaluators in Merlot. Considering this, the present work aims to analyze the agreement and the existence of associations between the users who performed evaluations inside the Merlot repository. For that, two case studies were carried out with data collected from Merlot. The first case study assessed the agreement between pairs of evaluators who rated the same set of resources and considering their preferred disciplines and area of expertise. The second case study measured the association between the average ratings of the resources given by evaluators coming from different categories of discipline. Results pointed out that evaluators present a great disagreement in the ratings they give for the same resource. Moreover, the average ratings of the resources considering the evaluators from different categories of discipline have shown no correlation. The results found here may be used to improve the mechanisms for quality assurance inside the repositories as well as the development of new forms of presenting the information related to the quality of the resources.

Keywords: Quality. Repository of objects and learning. Digital Resources. Agreement.

LISTA DE FIGURAS

Figura 1	Modelos de arquitetura	31
Figura 2	Ciclo de vida de um recurso digital com cinco fases	32
Figura 3	Ciclo de vida de um recurso digital com seis fases	33
Figura 4	Matriz de avaliação	39
Figura 5	Avaliação de qualidade do Merlot	44
Figura 6	Interface do Merlot	45
Figura 7	Força da correlação do valor de r_s	51
Figura 8	Esboço geral da metodologia	53
Figura 9	Modelo relacional da base de dados	55
Figura 10	Quantidade de materiais avaliados por pares de usuários avaliadores	57
Figura 11	Concordância geral por áreas preferencial iguais e diferentes	64
Figura 12	Concordância geral considerando as áreas de atuação profissional	71
Figura 13	Diagrama de dispersão da média das notas e as categorias	76
Figura 14	Interface de avaliação da qualidade do Merlot	79
Figura 15	Recomendação da interface	80

LISTA DE TABELAS

Tabela 1	Critérios para avaliação da qualidade LORI.....	38
Tabela 2	Exemplo da tabela dos pares de usuários avaliadores...	56
Tabela 3	Exemplo da tabela de disciplina de atuação preferencial	58
Tabela 4	Exemplo da tabela de subcategoria da disciplina de atuação preferencial	58
Tabela 5	Exemplo da tabela de atuação profissional	59
Tabela 6	Exemplo de dados utilizados.....	59
Tabela 7	Exemplo do agrupamento dos dados	60
Tabela 8	Concordância geral por áreas preferencial iguais e diferentes	65
Tabela 9	Concordância por áreas preferencial iguais	66
Tabela 10	Concordância por áreas preferencial diferentes	67
Tabela 11	Concordância por subcategorias iguais e diferentes	69
Tabela 12	Concordância por subcategorias iguais	70
Tabela 13	Concordância por subcategorias distintas	70
Tabela 14	Concordância considerando a mesma área de atuação profissional.....	72
Tabela 15	Concordância considerando áreas de atuação profissional distintas	73
Tabela 16	Associação entre as notas e a categoria do usuário e recurso.....	74

LISTA DE ABREVIATURAS E SIGLAS

ROA	Repositórios de Objetos de Aprendizagem.....	21
Merlot	The Multimedia Educational Resource for Learning and Online Teaching.....	22
PPGTIC	Pós-Graduação em Tecnologias da Informação e Comu- nicação.....	23
OA	Objeto de Aprendizagem.....	25
LOEM	Learning Object Evaluation Metric.....	26
PNL	Processamento de Linguagem Natural.....	26
LORI	Learning Object Review Instrument.....	27
BIOE	Banco Internacional de Objetos de Aprendizagem.....	30
ARIADNE	Alliance of Remote Instructional Authoring and Distri- bution Network for Europe.....	31
CAREO	Campus Alberta Repository of Educational Objects....	31
TIC	Tecnologias da Informação e Comunicação.....	34
IMS	Infrastructure Management Services.....	38
ISO	International Organization for Standardization.....	40
ICC	Intra-class Correlation.....	47

SUMÁRIO

1 INTRODUÇÃO	21
1.1 CONTEXTUALIZAÇÃO DO PROBLEMA E JUSTIFICATIVA DA PESQUISA	21
1.2 OBJETIVOS	23
1.2.1 Objetivo geral	23
1.2.2 Objetivos específicos	23
1.3 ADERÊNCIA DO OBJETO DE PESQUISA AO PPGTIC ..	23
1.4 METODOLOGIA	24
1.5 ORGANIZAÇÃO DO TRABALHO	24
2 REVISÃO DA LITERATURA	25
3 FUNDAMENTAÇÃO TEÓRICA	29
3.1 REPOSITÓRIO DE OBJETOS DE APRENDIZAGEM ...	29
3.1.1 Ciclo de vida dos objetos de aprendizagem	32
3.2 RECURSOS EDUCACIONAIS ABERTOS	34
3.3 AVALIAÇÃO DA QUALIDADE DENTRO DE REPOSITÓRIOS	37
3.4 MERLOT	42
3.5 CONCORDÂNCIA E CONFIABILIDADE ENTRE AVALIADORES	46
3.5.1 Intra-class Correlation	47
3.6 ANÁLISE DE CORRELAÇÃO E MEDIDAS DE ASSOCIAÇÃO	50
4 PROCEDIMENTOS METODOLÓGICOS	53
4.1 CONTEXTUALIZAÇÃO DOS DADOS	53
4.2 EXTRAÇÃO DOS DADOS	54
4.3 TRANSFORMAÇÃO DOS DADOS	55
4.3.1 Pré-processamento para o estudo de caso I	57
4.3.2 Pré-processamento dos dados para o estudo de caso II	59
4.4 CARREGAMENTO DOS DADOS	60
4.4.1 Carregamento dos dados para o Estudo de caso I ..	60
4.4.2 Carregamento dos dados para o estudo de caso II ..	61
5 RESULTADOS	63
5.1 ESTUDO DE CASO I - IDENTIFICAÇÃO DA CONCORDÂNCIA NAS AVALIAÇÕES DOS PARES DE USUÁRIOS AVALIADORES	63

5.1.1 Concordância considerando as disciplinas de atuação preferencial	63
5.1.2 Concordância considerando as subcategorias das disciplinas de atuação preferencial	67
5.1.3 Concordância considerando a atuação profissional ..	71
5.2 ESTUDO DE CASO II - CORRELAÇÃO ENTRE AS MÉDIAS DAS NOTAS DAS CATEGORIAS	73
5.2.0.1 Correlação entre as médias das notas de cada categoria e as categorias dos usuários	73
5.2.1 Recomendação da interface	76
6 CONCLUSÃO	83
REFERÊNCIAS	87
APÊNDICE A – Algoritmos utilizados para o pré-processamento das tabelas no estudo de caso I	99
APÊNDICE B – Algoritmos utilizados para o pré-processamento das tabelas no estudo de caso II	103

1 INTRODUÇÃO

1.1 CONTEXTUALIZAÇÃO DO PROBLEMA E JUSTIFICATIVA DA PESQUISA

Os repositórios de objetos de aprendizagem (ROA) são os sistemas que possibilitam o armazenamento (físico ou do metadado) dos recursos e que fornecem as funcionalidades essenciais para seu compartilhamento, manutenção e utilização. O aumento da quantidade de objetos de aprendizagem dentro dos diferentes repositórios existentes exige que essas plataformas adotem estratégias para a garantia da qualidade dos recursos armazenados. Considerando que os ROA, normalmente, atraem comunidades de usuários que compartilham dos mesmos interesses (HAN et al., 2008), uma das estratégias comumente utilizada para a garantia da qualidade dos recursos é aproveitar a colaboração e as interações dessas comunidades de prática no sentido de avaliar e opinar (impressões de uso e qualidade) sobre os diferentes recursos armazenados (CLEMENTS; PAWLOWSKI; MANOUSELIS, 2015).

Cechinel, Sánches-Alonso e García-Barriocanal afirmam que a garantia da qualidade de recursos em repositórios é uma tarefa complexa, pois ela incorpora diferentes aspectos e dimensões, como a qualidade do conteúdo e a usabilidade num âmbito amplo (CECHINEL; SÁNCHEZ-ALONSO; GARCÍA-BARRIOCANAL, 2011). Além disso, pode-se afirmar que a excelência dos recursos é de fundamental importância para o próprio repositório, sendo que estes repositórios aproveitam a colaboração e as interações dos seus usuários, oferecendo-lhes a possibilidade de avaliar e opinar sobre os diversos objetos neles armazenados (CLEMENTS; PAWLOWSKI; MANOUSELIS, 2015). Essa importância se dá, uma vez que as informações de qualidade (pontuações) são utilizadas pelos algoritmos de ranqueamento dessas plataformas para posicionar os recursos após uma busca realizada pelos usuários, pelos algoritmos de recomendação (quando os mesmos estão implementados), além de servir como referência para os usuários que estão buscando algum tipo de material específico (DOWNES, 2007).

Apesar dessa estratégia de avaliação pela comunidade já ser amplamente consolidada e adotada pelos principais ROA existentes (CLEMENTS; PAWLOWSKI; MANOUSELIS, 2015), a apresentação da qualidade de cada recurso acaba sendo realizada por meio de uma nota (pontuação) final única, que é calculada pela média de todas as notas fornecidas pela comunidade de usuários. Esse formato de apresentação omite a

distribuição das notas efetivamente fornecidas por toda a comunidade e esconde as informações referentes a como os diferentes tipos de grupos de usuários pontuaram um determinado recurso. Por exemplo, um determinado recurso educacional para a área de matemática pode ter sido muito bem avaliado por usuários que sejam professores de matemática, porém, mal avaliado por usuários de outras áreas. Essas nuances referentes às avaliações dos diferentes grupos simplesmente desaparecem na apresentação da qualidade dos mesmos dentro dos ROA. Assim, usando o exemplo acima, um usuário da área de matemática poderia não encontrar facilmente o OA bem avaliado na sua área porque outras áreas fizeram a sua pontuação baixar e, portanto, ser pessimamente ranqueado na apresentação dos resultados. Nessa linha, uma pesquisa anterior (CECHINEL; SÁNCHEZ-ALONSO, 2011) demonstrou que existe considerável discordância com relação a qualidade dos recursos entre a comunidade de usuários do Merlot (The Multimedia Educational Resource for Learning and Online Teaching) e a comunidade de especialistas avaliadores (*peer-reviewers*).

Nesse sentido, é importante avaliar o grau de concordância e associação nas avaliações para identificar quanto o instrumento de avaliação por nota é eficiente, e os resultados podem ser utilizados para propor novas formas de apresentação da avaliação de qualidade realizada pela comunidade de usuários, auxiliando, por exemplo, no projeto de interface dos ROA.

Neste contexto, o presente trabalho busca responder às seguintes questões de pesquisa:

- Q1:** Qual a concordância geral entre pares de usuários avaliadores de objetos de aprendizagem ?

- Q2:** Qual o grau concordância entre pares de usuários avaliadores considerando suas preferências por determinadas disciplinas, e subcategoria (subárea) da disciplina preferencial e suas áreas de atuação profissional?

- Q3:** É possível identificar alguma associação significativa entre as notas dadas pelos avaliadores e as categorias dos avaliadores e dos recursos?

1.2 OBJETIVOS

1.2.1 Objetivo geral

Analisar a concordância e a existência de correlação entre as classificações dadas por avaliadores da base de dados MERLOT.

1.2.2 Objetivos específicos

- Identificar e tratar os dados da base de dados Merlot.
- Realizar uma revisão de literatura sobre avaliação da qualidade em repositórios, concordância e confiabilidade entre avaliadores e correlação.
- Analisar como é realizado o processo de avaliação pelos usuários dos OA no repositório Merlot.

1.3 ADERÊNCIA DO OBJETO DE PESQUISA AO PPGTIC

O Programa de Pós-Graduação em Tecnologias da Informação e Comunicação (PPGTIC) tem como área de concentração Tecnologia e Inovação, dividida em três linhas de pesquisa, são elas: Tecnologia Computacional, Tecnologia Educacional e Tecnologia, Gestão e Inovação.

Este estudo foi desenvolvido na linha de pesquisa Tecnologia Computacional, pelo fato de desenvolver técnicas que auxiliam na resolução de problemas de natureza educacional. Os Capítulos 4 e 5 esclarecem o alinhamento desta pesquisa com os objetivos da linha de pesquisa Tecnologias Computacionais. Pode-se observar o uso de mecanismos computacionais desde a coleta dos dados, pré-processamento, cálculos, geração de gráficos e análises dos dados.

Mesmo a pesquisa tendo aderência com a linha computacional, busca resolver uma lacuna encontrada na área educacional. Os dados utilizados para realizar a pesquisa foram coletados do ROA Merlot que possuem uma abordagem educacional.

Analisando os trabalhos já realizados no PPGTIC, foram encontrados trabalhos que utilizaram tecnologias para resolver problemas na área educacional. Mesmo sendo trabalhos que utilizam dados educacionais, nenhum dos trabalhos propostos abordaram a concordância

e correlação nas avaliações fornecidas pelos usuários avaliadores utilizando recursos computacionais.

1.4 METODOLOGIA

Este trabalho é de caráter interdisciplinar, pois está contido nas áreas da educação e tecnologias da informação e comunicação. No que se refere à abordagem, neste estudo predomina a quantitativa, pois busca a medição de variáveis, estatísticas e comparações. Quanto a sua natureza, a pesquisa é classificada como básica, pois gera novos conhecimentos para o avanço dos ROA sem uma aplicação prática dos resultados obtidos. Em relação aos objetivos, esta pesquisa é considerada explicativa, pois busca obter respostas para problemas com base nos resultados dos estudos (testes) realizados.

1.5 ORGANIZAÇÃO DO TRABALHO

O restante do trabalho está organizado da seguinte forma. O Capítulo 2 apresenta o estado da arte do problema. O capítulo 3 apresenta a fundamentação teórica sobre repositório de objetos de aprendizagem, recursos educacionais abertos, avaliação da qualidade dentro de repositórios, descrição sobre o repositório Merlot, concordância e confiabilidade entre avaliadores e análise de correlação e medidas de associação. Os procedimentos metodológicos adotados na pesquisa são descritos no Capítulo 4, dividido em em quatro etapas (contextualização dos dados, extração dos dados, transformação dos dados e carregamento dos dados). Os resultados e discussão dos dados são apresentados no Capítulo 5. Por fim, são apresentadas as conclusões do trabalho e os trabalhos futuros.

2 REVISÃO DA LITERATURA

Este capítulo apresenta uma breve revisão bibliográfica acerca do tema estudado para o entendimento completo do contexto do trabalho. São apresentados distintos estudos envolvendo métodos de avaliação dos recursos do ROA para entender o que os autores buscam identificar nas pesquisas e se existe algum estudo relacionado com esta pesquisa.

Existem distintas definições para Objeto de aprendizagem (OA), na pesquisa foi adotada a definição de Wiley (2002) que define o OA como qualquer recurso digital que possui a funcionalidade de reutilização para apoio a aprendizagem.

Os OA estão destinados a se tornarem essenciais em ambientes de ensino devido ao aumento gradual dos usuários em busca de conteúdo na web, facilitando o acesso às informações e conhecimentos (MESEGUER, 2001). Esses OA são geralmente compartilhados em plataformas de ensino (PAWLOWSKI, 2007).

A disponibilização de OA de qualidade tem sido um dos principais potencializadores de uso efetivo dos OA, além da utilização desses recursos para facilitar a reutilização dos mesmos (WILEY, 2002). Vários estudos destacam a necessidade de encontrar novas formas de gerir a qualidade dos ROA de maneira sustentável (DOWNES, 2007).

Na literatura, constata-se que existem diferentes abordagens na concepção do que pode ser classificado como métodos de qualidade. A principal abordagem adotada pelos autores, é a definição de indicadores de qualidade utilizados para medir e acompanhar o desempenho dos OA.

O estudo de Atenas e Havemann (2014) enfatizou a gestão da qualidade nos ROA como solução para o provimento de material de qualidade aos usuários em potencial. Os autores elencaram 10 indicadores de qualidade destacando, entre eles, a avaliação por usuário, revisão por pares, uso de metadados padronizados, plataforma com múltiplas línguas, ferramentas de mídia social, detalhamento da licença e open source.

Clements, Pawlowski e Manouselis (2015) investigaram as abordagens de qualidade empregadas nos ROA, concentrando-se no funcionamento do processo de garantia da qualidade e na identificação dos principais critérios de uma abordagem da qualidade para que os desenvolvedores consigam padronizar as abordagens ao projetar um ROA ou melhorar um existente.

Já, Kay e Knaack (2008) desenvolveram o modelo LOEM (*Le-*

arning Object Evaluation Metric) com base em uma lista de critérios escolhidos por meio da revisão de literatura. Este modelo foi criado com objetivo de avaliar os recursos do ROA. Os autores identificaram que a interatividade, design, engajamento e usabilidade são critérios que se correlacionam e estão diretamente ligados a qualidade, aprendizado e engajamento. Os autores mencionam ainda que o modelo pode ser considerado confiável na avaliação dos objetos de aprendizagem.

Vários estudos analisam as características e funcionalidades do ROA que auxiliam na garantia da qualidade, como por exemplo as especificações de metadados (OCHOA; DUVAL, 2006; PARK; TOSAKA, 2010), reutilização (CLEMENTS; PAWLOWSKI, 2012), entre outros. Clements e Pawlowski (2012) analisa uma abordagem mais geral, como padrões e políticas para garantia de qualidade, além de avaliar a perspectiva de gestão da qualidade do sistema e serviços fornecidos. Kurilovas, Birieniene e Serikoviene (2011) identificam critérios de qualidade como, domínios tecnológicos, direitos pedagógicos e a transparência relacionada à propriedade intelectual dos recursos. Sampson e Zervas (2013) analisaram 14 ROA a fim de elencar novas características que sejam capazes de apoiar a gestão do conhecimento nas plataformas. Como resultado, foi proposta uma lista de funcionalidades que consegue suprir as necessidades da comunidade e cada atividade dos sistemas de gerenciamentos que foram divididos em três, sendo o sistema de componente, metadados e outros serviços de valor agregado pessoais.

Os processos de curadoria de REA também são assuntos bastante abordados, como no trabalho (VIEIRA, 2019) que compara o processo de curadoria de quatorze ROA buscando compreender o processo de curadoria e analisar os critérios utilizados nos distintos sistemas de curadoria dos Objetos de Aprendizagem. O autor conclui que o processo de curadoria dos objetos de aprendizagem ocorrem em dois momentos na inserção do recursos no ROA e pós-publicação. Para Cechinel (2017), o processo de curadoria pode ocorrer nas diferentes etapas do ciclo de vida dos objetos de aprendizagem que na maioria das vezes inicia na etapa de etiquetamento e oferta, onde garante a coleta de todas as informações para criação do metadado. Na etapa de seleção são disponibilizadas informações sobre a qualidade do recurso. Por fim, na etapa de uso e retenção é realizado o processo de avaliação da qualidade pela comunidade de usuário.

Pelaez et al. (2019) propõem um método de avaliação de REA através de uma abordagem técnica, utilizando tecnologias emergentes, como tecnologias semânticas e técnicas de Processamento de Linguagem Natural (PNL). Baseado no Modelo de Qualidade OCW, formado pelas

seguintes fases: 1 - desenvolvimento, distribuição e licenciamento, 2 - classificação acadêmica, 3 - apresentação do usuário, 4 - material de avaliação e suporte, 5 - requisitos tecnológicos e interoperabilidade, 6 - acessibilidade. O objetivo final é fornecer métricas de qualidade dos REA.

O estudo de Ochoa e Duvak (2009) mostrou uma análise diferente das mencionadas, os autores buscaram quantificar o crescimento dos ROA conforme o aumento de OA e usuários. No estudo de Dimitriadis et al. (2009) é discutido a questão do design dos REA considerando o quesito reutilização e os benefícios da definição de padrões simples. Recentemente Garzon (2018) definiu um modelo de análise da aprendizagem para incentivar os professores no desenvolvimento e avaliação de REA.

No trabalho de Vives (2007) é apresentado um método de avaliação com base nas interações dos usuários. Foi utilizado o instrumento de avaliação LORI (*Learning Object Review Instrument*) para identificar o grau de qualidade dos OA. Para desenvolver a métrica de previsão de qualidade dos recursos por interação dos usuários foi utilizada a análise de regressão linear. Os autores identificaram que existe uma relação alta entre as interações dos usuários e a qualidade dos objetos. Sendo, possível identificar recursos com baixa qualidade com base nas interações dos usuários.

Cechinel e Sánchez-Alonso (2011) analisam a qualidade do sistema do ROA Merlot, através das associações entre as avaliações fornecidas pelos usuários e os especialistas no MERLOT, com o objetivo de identificar se existe concordância nas avaliações. Este estudo é um dos poucos que aborda a questão da qualidade do ROA levando em consideração o usuário. Nenhum dos estudos citados apresentou modelos de abordagens considerando a garantia de qualidade do sistema ou serviços. Esse tipo de estudo é importante para melhorar a interpretação das informações e desenvolver métodos para projetar as abordagens de qualidade.

Diante, dos trabalhos descritos percebe-se a preocupação dos autores em criar métodos e abordagens para garantir a qualidade dos recursos dentro dos ROA. Mas, poucos trabalhos buscam identificar ou medir se as abordagens existentes são realmente eficientes. O trabalho de Cechinel e Sánchez-Alonso (2011), por exemplo, busca analisar a existência de associações das notas dadas pelos avaliadores especialistas (revisões por pares de especialistas) e pelos usuários dos OA. Então, podemos identificar que esta pesquisa é de fato a primeira a ser realizada. A qual busca verificar se existe concordância e correlação nas

notas dadas pelos usuários dos OA. Portanto, o resultado da pesquisa contribuirá com os processos já existente identificando se os métodos atuais são realmente confiáveis, ainda os resultados da pesquisa podem ser considerados nos critérios de melhorias a ser efetuados nos ROA.

3 FUNDAMENTAÇÃO TEÓRICA

3.1 REPOSITÓRIO DE OBJETOS DE APRENDIZAGEM

De acordo com Downes (2001), ROA são plataformas com diversas funcionalidades planejadas para facilitar o acesso a OA em diferentes formatos (fotos, vídeos, documentos, entre outros) permitindo o usuário procurar e utilizar os recursos disponibilizados. McGreal (2008) define ROA como um banco de dados digitais presentes na web com acesso a interface do usuário que disponibilizam diversos recursos, neste trabalho utiliza-se essa definição. A ideia básica do ROA (muitas vezes chamado de plataforma) é armazenar objetos de aprendizagem para posteriormente serem recuperados.

Essa plataforma constitui um ambiente web colaborativo, em que a comunidade (estudantes, professores e usuários comuns) cria recursos e participem da comunidade (MACIEL, 2018). Exemplos de ROA: Merlot, Ariadne, MEC RED, Portal do professor, entre outros.

McGreal (2008) classifica os repositórios à partir dos critérios que seguem:

1. **Tipo de armazenamento do OA:** os ROA existentes armazenam os conteúdos localmente, outros armazenam apenas os metadados com links que redirecionam para outro ROA chamados de referatório e os ROA que armazenam os conteúdos e os metadados.
2. **Definição da área:** atualmente existem repositórios que são considerados multidisciplinar por disponibilizar conteúdo de diversas áreas e com conteúdo de uma área específica (saúde, educação, matemática).
3. **Cursos completos:** ROA que oferecem conteúdo completo referente a alguma disciplina.
4. **Restrições de uso:** políticas de uso e compartilhamento de OA. Em alguns ROA é necessário apenas tornar-se membro, sem nenhum custo e possuir apenas uma inscrição (cursos, disciplinas), mas em outros é necessário arcar com alguma taxa.

O principal aspecto de um ROA de sucesso é a capacidade de interoperabilidade através de padrões e protocolos, além de fornecer um

conjunto de funcionalidades fundamental para disponibilizar o acesso aos objetos de aprendizagem oferecidos (HIGGS; MEREDITH; HAND, 2002). O trabalho de Mcgreal (2008) tem sido referência na identificação das principais funcionalidades de um ROA, pois elenca as funcionalidades básicas que um ROA deve possuir, que seguem:

- Pesquisar e buscar: a capacidade de buscar recursos na plataforma de maneira simples ou avançada.
- Solicitar: um OA armazenado no banco de dados.
- Recuperar: acessar recurso solicitado.
- Enviar: inserir recurso digital no ROA.
- Armazenar: inserir OA no repositório com identificador único para conseguir ser recuperado posteriormente.
- Reunir: conseguir acesso aos metadados de OA em outros repositórios para conseguir informações mais abrangentes nas buscas.
- Publicar: fornecer metadados dos recursos para outros ROA.
- Avaliar: fornecer avaliação da qualidade aos materiais inseridos na plataforma.

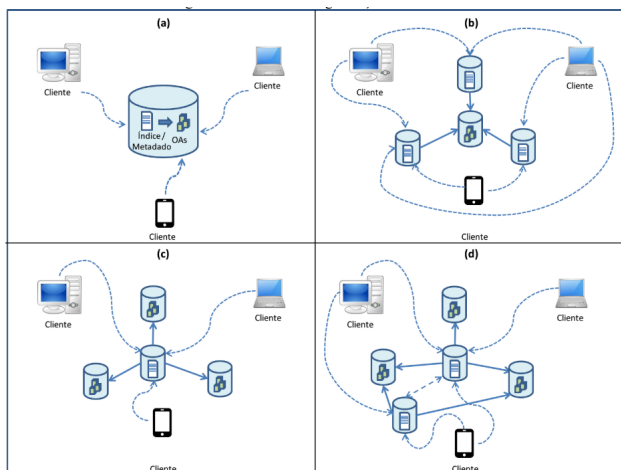
A partir da identificação das principais funcionalidades de um ROA, também é importante compreender como são organizados os ROA. Silveira, Omar e Mustaro (2007) apresentam quatro modelos de arquitetura da organização, como descrito a seguir e apresentado na Figura 1:

- **Objetos de aprendizagem centralizados e indexação de metadados centralizados** Figura 1(A) os objetos de aprendizagem e os metadados são armazenados em um servidor centralizado que é acessado por qualquer usuário. Como consequência, ocorrem altos custos de processamento e o desempenho para recuperar os recursos é lento. O repositório BIOE (Banco Internacional de Objetos de Aprendizagem) é baseado nesta arquitetura de organização.
- **Objetos de aprendizagem centralizados e indexação de metadados distribuídos** Figura 1(B) distribui os metadados e armazena os objetos de aprendizagem centralizados. Dessa forma,

consegue possuir diferentes formatos de metadados para referenciar os recursos. Nessa arquitetura é necessário utilizar uma boa infraestrutura para armazenar os recursos, mas muitas vezes não é possível armazenar os metadados.

- **Objetos de aprendizagem distribuídos e indexação de metadados centralizada** 1(C) os objetos de aprendizagem são armazenados em diferentes repositórios e os índices são mantidos centralizados. O Merlot e o ARIADNE são exemplos de repositórios que utilizam a arquitetura de objetos de aprendizagem distribuídos e uma indexação centralizada.
- **Objetos de aprendizagem distribuídos e indexação distribuída** 1(D) os repositórios, índices e metadados referem-se a múltiplos ROA. Essa infraestrutura permite que a indexação seja realizada por diferentes estratégias, índices e diferentes formatos de metadados. Essa alternativa é a mais confiável e robusta e também apresenta o maior desempenho, por outro lado, necessita de maior investimento. O repositório CAREO (Campus Alberta Repository of Educational Objects) é baseado nesta arquitetura de organização.

Figura 1 – Modelos de arquitetura



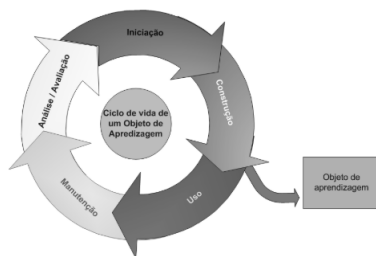
Fonte: (POTTKER, 2017)

A arquitetura está relacionada as tomadas de decisões realizadas de acordo com o funcionamento do ROA. Essas decisões são de natureza técnica, porém se relacionam com as demais áreas relacionadas a organização dos repositórios. A arquitetura básica de um ROA parte do princípio da necessidade de uma plataforma para armazenamento de conteúdo somado a um mecanismo de indexação e busca e uma interface para interação do usuário (BARRIT; ALDERMAN, 2004).

3.1.1 Ciclo de vida dos objetos de aprendizagem

O ciclo de vida de um recurso digital normalmente segue algumas etapas definidas. Como, por exemplo, o modelo apresentado na Figura 2, formado por 5 etapas. A primeira etapa é a **iniciação** onde é definida a atividade que será realizada, seja um novo recurso ou uma melhoria. A etapa seguinte é a **construção**, nesta etapa são desenvolvidas as funcionalidades para criar ou incorporar o recurso. Após a criação do recurso, é disponibilizado aos usuários o recurso para **uso**, assim, o próprio usuário consegue reportar erros encontrados durante o uso. Se os usuários identificarem erros, o recurso passa para a fase de **manutenção**, onde ocorre a correção destes erros. Por fim, é realizada a **avaliação** do recurso criado com os requisitos definidos na fase de iniciação (OLIVEIRA; NELSON; ISHITANI, 2007).

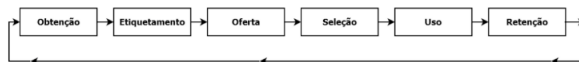
Figura 2 – Ciclo de vida de um recurso digital com cinco fases



Fonte: (OLIVEIRA; NELSON; ISHITANI, 2007)

Já o modelo de Collis e Strijker (2004), apresentado na Figura 3, mostra seis etapas do ciclo de vida dos objetos de aprendizagem.

Figura 3 – Ciclo de vida de um recurso digital com seis fases



Fonte: (CECHINEL, 2017) adaptado de (COLLIS; STRIJKER, 2004)

De acordo com Collis e Strijker (2004), a primeira fase do ciclo de vida de um recurso digital é a **obtenção**, onde o material é obtido em formato digital. Esta etapa consiste basicamente na criação do recurso digital, ou a utilização de template já existentes (CECHINEL, 2015). Após a criação do recurso, é necessário criar o metadado do mesmo para posteriormente facilitar na busca e documentação do recurso (COLLIS; STRIJKER, 2004). Segundo Cechinel (2015), a fase de **etiquetamento** consiste no fornecimento das informações necessárias que auxiliará na busca do recurso. O autor ainda informa que esse processo pode ser realizado de diversas maneiras, utilizando informações básicas ou completas do recurso, por meio de algum padrão de descrição ou sem uso de padrão. A terceira etapa do ciclo é a **oferta**, que disponibiliza ou publica um conjunto de objetos de aprendizagem na plataforma, de maneira que o usuário do recurso consiga acessar, baixar e utilizar (COLLIS; STRIJKER, 2004; CECHINEL, 2015; CARDINAELS, 2007). Após a publicação, os usuários conseguem realizar a busca pelo recurso. O processo de **seleção** é a etapa que os objetos de aprendizagem são pesquisados e selecionados nos repositórios. Cechinel (2015) explica que essa fase é considerada crítica, pois se um recurso nunca foi recuperado, o ciclo de vida do mesmo foi encerrado. Quando o usuário consegue selecionar o recurso desejável, é possível utilizá-lo. A fase de **uso** pode ser realizada de duas formas, a primeira é fazer uso do recurso digital de forma original. A segunda maneira é a adaptação do recurso selecionado, ou seja, o usuário consegue editar o recurso conforme o ambiente que deseja utilizá-lo (COLLIS; STRIJKER, 2004). A última fase é a **retenção**. Após o uso do recurso, é permitido reutilizar, revisar e remover o mesmo do repositório (CARDINAELS, 2007). Nesta fase, é observado se o recurso digital é útil na plataforma. É importante realizar o acompanhamento das ferramentas de qualidade para identificar se o recurso criado está sendo utilizado e verificar se é necessário realizar alguma modificação ou excluir do repositório (COLLIS; STRIJKER, 2004).

3.2 RECURSOS EDUCACIONAIS ABERTOS

A disseminação das Tecnologias da Informação e Comunicação (TIC) no setor educacional, trouxe grandes mudanças nos procedimentos educacionais, principalmente no ensino superior (PELAEZ et al., 2019). Os usuários de todo o mundo estão criando conjuntos de recursos digitais abertos e gratuitos para uso global, além de ser uma forma de disseminar o conhecimento e facilitar o acesso a materiais, está se tornando uma maneira de fornecimento de ensino (YUAN et al., 2008). Para alcançar esse público são definidas três estratégias principais: incentivar usuários a participarem ativamente do processo da educação aberta, os REA devem ser compartilhados através de licenças abertas permitindo que o usuário fique assegurado do uso e reúso do material, além de seguir normas de publicação facilitando o compartilhamento e instituições governamentais devem ter como objetivo principal a educação aberta (YUAN et al., 2008). Com o surgimento dos REA, a proposta de um novo modelo de ensino ampliou a educação aberta garantindo uma nova oportunidade de aprendizado com menos recursos e democratização da educação (WILLIAM; FOUNDATION, 2013). Os REA são considerados por Hylén (2006) materiais digitais oferecidos gratuitamente e abertos com a finalidade de reúso no ensino e aprendizagem, como por exemplo a iniciativa Khan Academy, o MIT OpenCourseware e a Open Course Library from Washington (WILEY; GREEN; SOARES, 2012). Milhares de REA são disponibilizados nos ROA diariamente com a finalidade de reutilização, publicação e compartilhamento nas comunidades, além de oferecer recursos para criação de REA colaborativamente (DIMITRIADIS et al., 2009).

Os materiais inseridos nos ROA possuem algum tipo de licença de propriedade intelectual que permite o uso gratuito ou a reutilização ou são de domínio público, esses objetos de aprendizagem possuem diversos formatos sejam cursos completos, materiais, módulos, livros didáticos, vídeos, software e quaisquer outras ferramentas, materiais ou técnicas usadas para apoiar o acesso ao conhecimento (TUOMI, 2013). Nos REA o usuário consegue realizar quatro tipos de atividades, sendo (WILEY; GREEN; SOARES, 2012):

- **Revisar:** adaptar os materiais disponibilizados conforme as necessidades.
- **Reutilizar:** utilizar o recurso de diferentes maneiras e contexto.
- **Combinar:** combinar diferentes REA para criar novos materiais.

- **Redistribuir:** compartilhar REA com a comunidade

Garantir e controlar a qualidade desses objetos de aprendizagem é uma tarefa difícil, sendo essencial para criar ferramentas adequadas, manter o REA disponível e fornecer conteúdo com informações de qualidade aos usuários em potencial (DOWNES, 2007). De acordo com Hurtado (2019) o maior obstáculo que os ROA vem enfrentando é o controle de qualidade dos materiais disponibilizados, a falta de estratégias e a dificuldade de encontrar REA com conteúdo adequado afeta os usuários diretamente. A avaliação da qualidade dos REA é voltada às características que cada recurso apresenta, sendo observadas em um contexto de ensino e aprendizagem (GARZON, 2018).

Atualmente alguns ROA que fornecem recursos abertos estão adotando algumas abordagens de gerenciamento da qualidade dos objetos de aprendizagem disponibilizados, que são classificadas em três tipos, sendo reputação, revisão por pares e interação do usuário (HYLÉN, 2006). Segundo Yuan et al. (2008) a **Abordagem institucional** refere-se a confiança do usuário na reputação da instituição, acreditando que os materiais no site são de boa qualidade, como as iniciativas OCW e a iniciativa OpenLearn da Universidade Aberta do Reino Unido. **Abordagem de revisão por pares** um método de garantia de qualidade muito utilizado nas universidades. **Abordagem por usuário** a prática com menor custo, onde o usuário pode avaliar se um recurso de aprendizado é de alta qualidade, útil ou bom em qualquer outro aspecto.

A seguir, são apresentados alguns programas e projetos com iniciativas de REA:

Connexions - um repositório que permite reunir conteúdo, comunidades e software em um ambiente de ensino e aprendizagem (YUAN et al., 2008). Pelo fato do repositório disponibilizar um conjunto de ferramentas de software livre para publicação e colaboração entre usuários, todos os recursos disponibilizados são criados pela própria comunidade de usuário (OCHOA; DUVAL, 2009) e todo material criado é disponibilizado na plataforma de forma livre¹.

MIT OpenCourseWare - tem como objetivo fornecer uma política e modelo de distribuição de cursos gratuitamente (DOWNES, 2007). Todos os cursos disponibilizados são para uso reúsio sem nenhum custo (CLEMENTS, 2016), para que os usuários consigam utilizar para o ensino, como ferramenta de desenvolvimento, planejamento e criação de cursos ².

¹<https://cnx.org/>

²<https://ocw.mit.edu/>

Khan Academy - uma organização sem fins lucrativos com a missão de oferecer uma educação gratuita de alta qualidade para qualquer pessoa, em qualquer lugar. Oferecem exercícios, vídeos de instrução e um painel de aprendizado personalizado que habilita os estudantes a aprenderem no seu próprio ritmo dentro e fora da sala de aula. Os recursos são criados por especialistas, mas permitem a reutilização, de modo que os professores conseguem identificar as dificuldades de compreensão de seus alunos, personalizar instruções e atender às necessidades de cada um deles ³.

Esses ROA abertos são mantidos por meio de financiamento, atualmente existem diversos tipos de fontes de financiamento para iniciativa aberta de recursos educacionais. De acordo com Downes (2001) os principais modelos são os seguintes:

- Doação - Geralmente o projeto é mantido por meio dos juros obtidos pelo fundo. Esse fundo é regido por fundações, onde oferece um valor básico para sustentar o projeto.
- Associação - organizações contribuem com um determinado valor. Seja um valor único, uma contribuição ou um valor mensal.
- Doação da comunidade: a comunidade de usuário contribui com um determinado valor quando o projeto solicita ajuda de custos. Um exemplo é o Wikipédia, mas inúmeros projetos solicitam também é colaboração com conteúdo ou código.
- Conversão - ocorre quando oferece alguns serviços ou produto gratuitamente e consegue-se converter o consumidor em um cliente.
- Pagamento de contribuinte - os contribuintes pagam pela manutenção do sistema e o provedor disponibiliza a contribuição gratuitamente.
- Patrocínio - várias empresas se unem para apoiar um projeto, geralmente essas empresas mantêm uma parceria com as instituições de ensino.
- Institucional - uma instituição assumirá a responsabilidade por uma iniciativa aberta.
- Governamental - o financiamento para projetos é diretamente proveniente de agências governamentais.

³<https://www.khanacademy.org/>

Os REA estão cada vez mais presentes na web e apresentam uma série de benefícios para educação. No trabalho de Yuan et al. (2008) é elencado alguns benefícios que os REA promovem, sendo:

1. Compartilhamento de conhecimento, permite que a formação de comunidades de usuários em busca de novos conhecimentos.
2. A possibilidade de compartilhamento e a reutilização de forma gratuita dos recursos.
3. Criação de conteúdo compartilhada, ou seja, os usuários conseguem colaborar com outros membros na criação de materiais. Com isso, a qualidade pode ser melhorada e o custo da criação de conteúdo é reduzido pelo compartilhamento e reutilização.
4. Facilidade de acesso atrai diversos usuários.
5. O compartilhamento aberto dos REA, incentiva a melhoria da qualidade dos ROA, a inovação e o reúso. Além das plataformas conseguirem manter materiais de qualidade e facilitar na criação de novos recursos.

3.3 AVALIAÇÃO DA QUALIDADE DENTRO DE REPOSITÓRIOS

Os objetos de aprendizagem estão demonstrando um papel importante no processo de ensino e na aprendizagem, devido a criação de padrões de e-learning e o surgimento dos ROA (LI et al., 2015) . Os ROA fornecem um rico ambiente de aprendizado multimídia e oferecem um novo paradigma de aprendizado. Embora, a criação de recursos de aprendizado on-line variam muito e os procedimentos ou diretrizes de controle de qualidade necessitam serem eficientes e apresentarem os requisitos de confiabilidade e qualidade, muitos objetos de aprendizagem não seguem tal formato (VARGO et al., 2003).

Os ROA apresentam diversos problemas que garantem a qualidade dos dados apresentados (PAWLOWSKI, 2007), por falta de informação de conteúdo e especificação da avaliação de qualidade. Neste caso, é necessário além de garantir a qualidade dos objetos de aprendizagem nos repositórios, deve-se também verificar a qualidade das funcionalidades fornecidas pelos repositórios aos usuários (CHARLESWORTH et al., 2008). Para isso foram desenvolvidas diversas abordagens de garantia de qualidade. Como por exemplo o instrumento de medição quantitativa da qualidade de objetos de aprendizagem LORI (NESBIT et al.,

2002), criado para definir métricas para medir a qualidade pedagógica dos OA. De acordo com Gordillo, Barra e Quemada (2014) os revisores podem usar o LORI para avaliar os OA conforme os critérios apresentados na Tabela 1, além de gerar uma métrica geral da qualidade de todos os critérios. O autor explica que a avaliação é realizada através de comentários e classificações em uma escala de 5 pontos, onde 1 significa baixo e 5 alto. Os revisores podem deixar critérios sem avaliação escolhendo a opção "Não aplicável".

Tabela 1 – Critérios para avaliação da qualidade LORI

Critério	Descrição
Qualidade de conteúdo	veracidade, precisão, apresentação equilibrada de ideias e nível de detalhe apropriado.
Alinhamento com o objetivo de aprendizagem	alinhamento entre objetivos de aprendizagem, avaliação de atividades e características do aluno
Feedback e adaptação	conteúdo adaptável ou feedback orientado por informações de aprendizagem diferenciais ou modelagem de alunos.
Motivação	capacidade de motivar e interessar populações identificadas de alunos
Design de apresentação	projeto de informações visuais e auditivas para aprendizado aprimorado e processamento mental eficiente
Usabilidade de interação	facilidade de navegação, previsibilidade da interface do e segue os padrões da IMS <i>Guidelines for Accessible Learning Applications</i> .
Acessibilidade	acomoda as necessidades dos aprendizes com deficiência acomodar pessoas com deficiência e mobilidade
Reusabilidade	capacidade de usar em diferentes contextos de aprendizagem e com alunos de diferentes origens
Conformidade com padrões	adesão às normas e especificações internacionais

Fonte: (GORDILLO; BARRA; QUEMADA, 2014), tradução nossa

A garantia da qualidade de objetos de aprendizagem pode envolver diferentes opiniões nos recursos ou possuir apenas um tipo de controle interno de qualidade da plataforma. Desse modo, um outro modelo utilizado é a matriz apresentada por Hylen (2006), que mostra quatro eixos associados as principais características de avaliação.

As características de avaliação estão distribuídas em quatro quadrante principais, onde define as principais abordagens de avaliações existentes nas plataformas. O primeiro quadrante apresentado na Figura 4 é o *Peer review* essa abordagem realiza revisões de recursos por pares especialistas, onde existe uma segurança pelo reconhecido conhecimento desses especialistas e pelo conjunto de critérios de avaliação que são pré-estabelecidos no processo de avaliação. A Segunda abordagem *User comments, user rating* é largamente adotada pelos principais repositórios, na qual permiti que os usuários decidam, por meio de pontuações e comentários se a qualidade do recurso é boa ou não. A terceira abordagem *Word of mouth* é utilizada por plataformas que possuem uma boa reputação, assim consegue convercer os usuários que seus recursos são de boa qualidade. Por fim, *Internal quality proce-*

dures a recomendação dos objetos de aprendizagem é realizada de boca a boca.

O eixo horizontal abrange a avaliação descentralizada e centralizada e o eixo vertical refere-se a avaliação aberta e fechada, Figura 4.

Figura 4 – Matriz de avaliação



Fonte: (HYLÉN, 2006)

A avaliação aberta e fechada refere-se a transparência durante a avaliação. O processo de avaliação é considerado aberto quando as avaliações, recursos e critérios são totalmente explícitos a qualquer tipo de usuário, no entanto, a avaliação fechada refere-se ao processo contrário do citado (VIEIRA, 2019). De acordo com Vieira (2019), o segundo eixo considera o controle durante o processo da curadoria. O processo de curadoria refere-se basicamente ao processo de selecionar, avaliar, organizar, administrar e comparar os recursos e funcionalidades dos objetos de aprendizagem de maneira que eles possam ser reutilizados e compartilhados com a comunidade de usuários (CECHINEL, 2017). A avaliação é considerada centralizada quando o processo de avaliação está somente no controle da plataforma. É considerado o processo de avaliação descentralizado quando o controle da qualidade está diretamente relacionado ao usuário através de comentários e notas.

As abordagens de avaliação da qualidade utilizadas nos ROA são classificadas por Clements, Pawlowski e Manouselis (2015) em três categorias. A primeira categoria abrange as normas de **abordagem genérica da qualidade** que fornecem um padrão ou procedimento in-

dependente do domínio como, por exemplo, o ISO 9000, o que garante que as organizações consigam um selo de aprovação por meio de avaliação externa (EHLERS; PAWLOWSKI, 2006). São amplamente utilizados, mas o esforço para adaptar essa abordagem é muito alto.

A segunda categoria envolve as **abordagens de qualidade específica**, são abordagens de qualidade que lidam com o domínio da aprendizagem, educação e formação, especificamente e-learning (PAWLOWSKI, 2007). Fornecem procedimentos de garantia de qualidade para o domínio de Aprendizado aprimorado por tecnologia. Abordagens específicas visam alcançar a alta qualidade nos ROA. Para evitar os grandes esforços de adaptação, as abordagens específicas para o campo da aprendizagem, educação e treinamento foram desenvolvidas

A terceira categoria é a que utiliza algum **instrumento específico de qualidade** em que recursos são avaliados pela comunidade de usuários (ATENAS; HAVEMANN, 2014). É considerada a categoria que mais se popularizou entre os ROA para assegurar a qualidade de seus recursos (CLEMENTS; PAWLOWSKI; MANOUSELIS, 2015), sendo que as avaliações podem ser realizadas por meio de critérios de qualidade previamente definidos e estruturados. Na classificação de avaliação de objetos de aprendizagem em ROA proposta por Hilén (2006) e apresentada na Figura 4, essa abordagem é considerada aberta (todas as pontuações e comentários são possíveis de serem visualizados no ROA) e descentralizada (o gerenciamento do ROA não é realizado diretamente, e todas as avaliações são realizadas de maneira livre). Os principais instrumentos específicos de qualidade utilizados nos ROA atualmente são:

- **Revisão por pares de especialistas:** Cechinel (2015) explica que a revisão por pares de especialistas é o processo de avaliar um trabalho de forma crítica por especialistas com domínio na área. Sendo uma atividade demorada que requer esforço por parte da organização (NEVEN; DUVAL, 2002). Esse processo de avaliação é utilizado geralmente em processo de publicação de artigos, onde especialistas revisam o trabalho como forma de avaliar a qualidade do material submetido (BENOS et al., 2007) e identificar possíveis alterações que o trabalho poderá sofrer para uma sucessiva aprovação (OCHOA; DUVAL, 2009). Uma abordagem de qualidade adotada por muitos ROA (WINDLE et al., 2010), por exemplo, MERLOT, ARIADNE e OER Commons.
- **Revisão por usuários:** a comunidade de usuários é o principal meio de garantir a qualidade (LARSEN, K; LANCIN, 2005). Os

usuários participam do processo de avaliação da qualidade geralmente por meio de avaliações, classificações e recomendações (CLEMENTS; PAWLOWSKI, 2012). O Edmodo Spotlight e o Merlot são exemplos de ROA que permitem o usuário avaliar os materiais por meio de postagens e pontuações.

- **Sistema de recomendação:** são sistemas que podem guiar automaticamente os usuários fornecendo recomendações de objetos de seus interesses anteriores, gostos semelhantes com base as variáveis acessadas na plataforma (GOLDBERG et al., 1992; BURKE, 2002; MANOUSELIS; SAMPSON, 2004). O Elera adota o sistema de recomendação de materiais que permite ser recomendado por pontuações e popularidade (NESBIT; LI, 2004). Já o Connexions possui plugins associados a redes sociais, permitindo que o usuário recomende materiais através de suas redes sociais (CECHINEL, 2015).
- **Favoritos:** é uma funcionalidade que permite os usuários criarem coleções e organizar os materiais de acordo com seus interesses (MINGUILLÓN; RODRÍGUEZ; CONESA, 2010; SAMPSON; ZERVAS, 2013). O Currículo+ é um exemplo de ROA que fornece essa funcionalidade aos usuários.
- **Política de qualidade:** são normas criadas para permitir o compartilhamento legalmente do conhecimento (MACEDO, 2010). Um exemplo é a Creative Commons é uma organização sem fins lucrativos que permite o compartilhamento e uso da criatividade e do conhecimento através de instrumentos jurídicos gratuitos (GARLICK, 2005).

Exemplos de repositório que utilizam essas abordagens são: OpenStat Cnx⁴, Common Sense Education⁵, Edmodo Spotlight⁶, o Merlot⁷ e a Plataforma MEC de objetos de aprendizagem⁸.

No caso específico do repositório Merlot, além da avaliação por parte da comunidade de usuários (por meio de comentários e pontuações), o ROA também utiliza um modelo de revisão por pares especialistas, além de permitir a organização dos recursos em coleções pessoais e conceder um prêmio (selo) de qualidade para os recursos mais bem

⁴<https://cnx.org/browse>

⁵<https://www.commonsense.org/education>

⁶<https://spotlight.edmodo.com>

⁷<https://www.merlot.org>

⁸<https://plataformaintegrada.mec.gov.br/>

pontuados e comentados pelos especialistas e pela comunidade (Merlot classic awards)(KUTLUCA, 2010). Todas essas estratégias são complementares e utilizadas pelo ROA para o ranqueamento dos recursos nas buscas realizadas pelos usuários.

O presente trabalho lida diretamente com as avaliações dos objetos de aprendizagem dentro do repositório Merlot e que foram realizadas pela comunidade de usuários, portanto pertencendo a terceira categoria de estratégia de avaliação mencionada, e sendo classificada como uma abordagem de avaliação aberta e descentralizada.

3.4 MERLOT

O Merlot é uma iniciativa internacional que busca fornecer um ambiente gratuito, de fácil acesso, que conecta diferentes tipos de pessoas (professores, funcionários, alunos, entre outros) a materiais de aprendizagem online (SHEA; MCCALL; OZDOGRU, 2006). O projeto da plataforma iniciou em 1997, quando o Centro para Aprendizagem Distribuída da Universidade do Estado da Califórnia desenvolveu e forneceu acesso gratuito ao Merlot. Foram desenvolvidas ferramentas para ajudar na formação da comunidade, fornecendo conhecimentos básicos por meio de materiais de aprendizagem. Em 1999, reconheceram que uma iniciativa de cooperação era significativa para expandir as coleções dos materiais, conduzir as revisões dos objetos de aprendizagem e adicionar novas funcionalidades de aprendizagem para os alunos. Em seguida, foi necessário desenvolver um processo de avaliação. Então professores foram contratados para criar padrões de avaliação e processos de revisão por pares para material de ensino online. Logo após, diversas instituições foram convidados a ingressar no projeto tornando-se parceiros do Merlot.

O Merlot também possui uma comunidade de voluntário que tem o objetivo de colaborar com a criação de novas coleções e adicionar atribuições e comentários. Além disso, ajudam na divulgação da plataforma em eventos e trabalham com a equipe para desenvolver políticas e práticas que governam as operações . Ainda, possuem a comunidade destinada as disciplinas, apoio acadêmico e de parceiros.

O repositório é formado por mais de 40.000 materiais pertencentes a 19 categorias diferentes (Artes, Negócios, Tecnologias, entre outras). Os materiais inseridos no Merlot são fornecidos pelos membros cadastrados na plataforma. Todos os serviços oferecidos pela plataforma são totalmente gratuitos. O Merlot é classificado como refera-

tório (CECHINEL, 2015) devido a forma de armazenamento dos dados. A plataforma mantém apenas os registros dos metadados de cada material e cria um link para cada material que redireciona para o próprio site.

O Merlot possui algumas ferramentas de avaliações tanto por especialistas quanto pelos usuários. A avaliação de qualidade por especialista é assegurada pós-publicação (CAFOLLA, 2002) e possuem 3 etapas para concluir uma revisão por pares:

1. **Desenvolvimento de padrões de avaliação** por meio de fornecimento de treinamento para os revisores com a finalidade de compreender os padrões e processos das avaliações, buscando garantir a qualidade do conteúdo, eficácia no potencial da ferramenta de ensino e facilidade de uso.
2. **Realização de avaliações** primeiro é realizada uma triagem através da classificação de importância para selecionar os materiais para revisão dos pares especialistas. Por fim, é atribuído a cada par de revisor um material. Para a avaliação, são utilizados procedimentos pré-definidos por formulários e padrões de avaliações.
3. **Relatórios de avaliações** é criado um relatório de revisão por pares que envia a revisão por pares ao autor do material para obter feedback e permissão para postar a revisão.

MERLOT tem mais de 20 conselhos editoriais, cada um dirigido por um editor e inclui um grupo de revisores pares. Cada Conselho Editorial tem um Editor que faz a triagem de cada material em sua disciplina e o designa a um revisor em sua disciplina.

A avaliação feita pela comunidade de usuários é fornecida por comentários e notas, variando de 1 à 5. Como o Merlot possui dois grupos (usuários e especialistas) muito bem definidos que fornecem as notas, pode ocorrer de apresentar grandes divergências de opinião sobre um mesmo assunto (CECHINEL; SÁNCHEZ-ALONSO, 2011). A Figura 5 mostra como é apresentada a informação da qualidade ao usuário, na qual estão separadas as notas dadas pelos pares especialistas e os usuários.

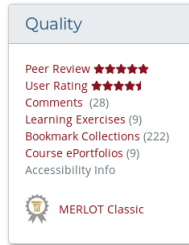


Figura 5 – Avaliação de qualidade do Merlot

O Merlot também possui a funcionalidade de marcação de favoritos, onde o usuário consegue criar sua própria coleção de materiais em seu perfil e compartilhar com a comunidade.

A interface do usuário do Merlot é bastante intuitiva, sendo distribuídas as funcionalidades em seis tópicos (Browse, Add, Communities, Partner Benefits, News & Info e About Merlot). Além da funcionalidade de ajuda, cadastro e acesso restrito aos membros. Cada tópico possui uma série de recursos que pode ser mostrada na Figura 6.

- **Browse:** o usuário pode realizar buscas sobre materiais, membros, exercícios, coleção de favoritos, cursos e portfólios, par de revisão, entre outros. O usuário consegue filtrar a busca por disciplinas, tipo de usuário, tipo de afiliação e outros filtros mais específicos. Cada usuário ou material possui uma descrição detalhada sobre o mesmo, facilitando a identificação do conteúdo e acesso ao material desejado.
- **Add:** é a barra de tarefa que permite o usuário adicionar algum recurso no ROA. Adicionar um material ao MERLOT, criar materiais com outro usuário e criar um portfólio de curso. Para acessar essas funcionalidades o usuário precisa ser membro do Merlot, ou seja, ser cadastrado na plataforma.
- **Communities:** o usuário pode fazer parte das comunidades de disciplina acadêmica, suporte acadêmico e parceiros. A comunidade de disciplina acadêmica e suporte são divididas de acordo com a disciplina, no qual são discutidos assuntos e problemas relacionados a categoria da disciplina. Já a comunidade de parceiros que hoje são formadas por 48 grupos, disponibilizam informações sobre suas parcerias formadas com o ROA.
- **Partner Benefits:** apresenta todas as informações de como tornar um parceiro do Merlot.

- News & Info: o usuário consegue navegar nas novidades e informações que o Merlot vem apresentando através das redes sociais, artigos e notícias. O tópico é subdividido em novidades do MERLOT, nossa conferência, programas e projetos Merlot, centro de mídia, vozes Merlot, blog e redes sociais (facebook, twitter e youtube).
- About Merlot: o usuário consegue acessar todas informações detalhadas sobre cada funcionalidade disponibilizada no Merlot. As informações são destinadas ao usuário que tem interesse de entender melhor como funciona o repositório. Sendo possível pesquisa sobre o repositório, coleções, tecnologias, centro de mídia, informações de como se tornar um parceiro, informações de como é realizada a revisão por pares, políticas e práticas, sobre a conferência e acessibilidade.

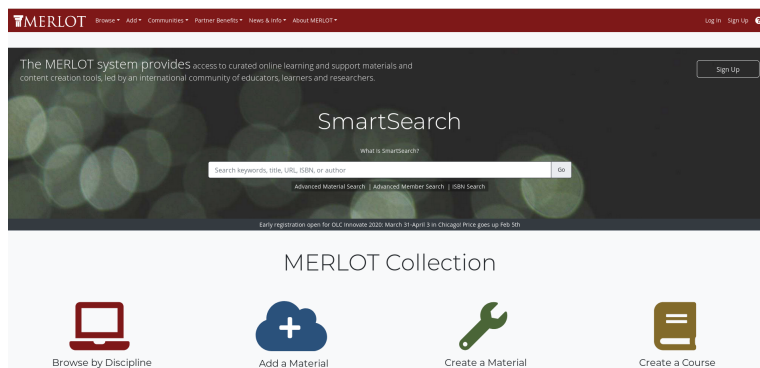


Figura 6 – Interface do Merlot

O Merlot também disponibiliza a funcionalidade de seleção de idioma, ou seja, o usuário consegue determinar em qual idioma deseja que as informações do repositório sejam apresentadas.

A Funcionalidade de busca e pesquisa atendem a qualquer necessidade. A pesquisa de materiais pode ser realizada por palavras-chave, título, URL, descrição, comunidade, categoria, idioma, nível, tipo de material, formato técnico, audiência, autor, licença e várias outras opções.

Para cada objeto de aprendizagem existe uma página inicial que descreve todas as informações e características, além de permitir a inclusão de comentários e avaliações por parte do usuário.

3.5 CONCORDÂNCIA E CONFIABILIDADE ENTRE AVALIADORES

Existem, basicamente, dois termos principais utilizados para trabalhar com o cálculo da consistência entre as pontuações dadas por diferentes avaliadores (às vezes chamados de juízes), sendo eles a *concordância* e a *confiabilidade*. Muitas vezes esses termos são utilizados de maneira intercambiável e generalizada, no entanto, confiabilidade e concordância de avaliadores possuem algumas diferenças importantes de serem mencionadas (SHWETA; BAJPAI; CHATURVEDI, 2015; SHOUKRI, 2010; MATOS, 2014).

A *confiabilidade* refere-se à medida da consistência entre notas fornecidas pelos avaliadores, seja na classificação ou na posição relativa das classificações de desempenho/notas fornecidas (SHWETA; BAJPAI; CHATURVEDI, 2015; VET et al., 2006). Por outro lado, a *concordância* é a medida em que dois ou mais avaliadores que usam a mesma escala de classificação e avaliam os mesmos materiais fornecem as mesmas notas, ou seja a concordância mede a frequência com que os avaliadores atribuem exatamente a mesma classificação/pontuação para os recursos/objetos pontuados(avaliados) (TINSLEY; WEISS, 2000; VET, 2014). Por exemplo, se dois avaliadores (Avaliador A e o Avaliador B) fornecerem as respectivas notas para quatro materiais: *Avaliador A* (1, 2, 3, 4) e o *Avaliador B* (2, 3, 4, 5). Os valores em parênteses referem-se a simulação das notas fornecidas por dois avaliadores. Temos como resultado que os avaliadores obtiveram uma confiabilidade alta e uma concordância baixa. Os avaliadores concordam com a qualidade do material, pois ambos forneceram notas que aumentavam gradativamente. O material 1 recebeu a nota mais baixa por ambos e o material 4 a mais alta, mas em nenhum momento eles concordam de maneira absoluta, apenas relativamente. Mudando o cenário de avaliação para o *Avaliador A* fornecendo as seguintes notas (1, 2, 3, 4) e o *Avaliador B* (1, 2, 3, 4), mostra que tanto a confiabilidade e a concordância são excelentes, ou seja, os avaliadores deram as mesmas notas e seguiram o mesmo padrão de avaliação (GRAHAM; MILANOWSKI; MILLER, 2012).

Uma das medidas de concordância bastante conhecida e utilizada para variáveis binárias é o coeficiente Kappa de Cohen que permite a avaliação simultânea de no máximo dois avaliadores (MATOS, 2014). O coeficiente Kappa é um método estatístico que considera a concordância ao acaso, ou seja, calcula a proporção de concordância entre os juízes após subtrair a proporção de concordância devido ao acaso (KRAEMER, 1980; BLOCH; KRAEMER, 1989). É um método muito utilizado,

mas apresenta muitas limitações pelo fato de não considerar aspectos importantes dos dados, então deve ser combinado com outras medidas de concordância (PERROCA; GAIDZINSKI, 2003).

A correlação intraclasse (ICC, por seu acrônimo em inglês) é uma outra medida de concordância, mas que também pode ser utilizada para medir a confiabilidade. Esta medida permite que o cálculo de concordância seja feito entre avaliadores para dados quantitativos, já a comparação é feita a partir da variabilidade de diferentes classificações do mesmo objeto com a variação total em todas as classificações dos objetos (KOO; LI, 2016). Existem ainda outros métodos para o cálculo da consistência entre avaliadores, como o Kappa Ponderado (concordância) e o Alpha de Krippendorff (confiabilidade e concordância) (MATOS, 2014).

3.5.1 Intra-class Correlation

No presente trabalho utilizou-se o ICC para medir a concordância nas avaliações entre pares de usuários avaliadores, pelo fato de atender todos os critérios necessários para a análise. Os critérios definidos foram: calcular o coeficiente para cada par de usuário avaliadores que possuem conjuntos de objetos comumente avaliados, a escolha dos pares de usuários avaliadores não é aleatória (e sim um consenso), considera uma amostra grande de avaliadores e o conjunto de dados analisados são classificados pelos mesmos pares de usuários avaliadores. Não foi calculada o grau de confiabilidade, pois, o interesse era de medir a frequência com que os avaliadores atribuem exatamente a mesma pontuação para os recursos.

Para calcular o ICC existem diferentes fórmulas que podem ocasionar resultados distintos quando aplicado ao mesmo conjunto de dados, pois pode ocorrer interpretações diferentes pelos pesquisadores. Para escolher a fórmula correta para os tipo de dados que coletamos, foi considerado o **Modelo** (o que queremos calcular), o **Tipo** (se é um único avaliador / média das medidas de k avaliadores / medidas) e a **Definição** (se busca calcular a concordância ou a confiabilidade) (KOO; LI, 2016). A escolha da fórmula para calcular a concordância geralmente é feita com base em quatro perguntas (KOO; LI, 2016).

1. Será utilizado o mesmo conjunto de dados para todos os avaliadores (define o modelo)?
2. A amostra foi selecionada aleatoriamente ou foi selecionada uma

amostra específica (define o modelo)?

3. Será calculado o coeficiente para um único avaliador ou o valor médio de vários avaliadores (define o tipo)?
4. O interesse é em calcular a confiabilidade ou a concordância (seleciona a definição)?

Os **Modelos** são divididos em três tipos:

1. **Efeito aleatório de uma via/unidirecional** (*one-way random*) - cada objeto é avaliado por um conjunto distinto de avaliadores, escolhidos aleatoriamente no conjunto de dados. (KOO; LI, 2016).
2. **Efeito aleatório de duas vias/bidirecional** (*two-way random*): - os avaliadores são selecionados aleatoriamente de acordo com suas características semelhantes (BARTKO, 1966).
3. **Efeito misto bidirecional** (*two-way mixed*) - este modelo é utilizado quando os avaliadores selecionados são os únicos avaliadores de interesse (LI et al., 2015).

O cálculo do ICC também é dividido em dois **Tipos** de mensuração, sendo que o primeiro tipo utiliza o valor médio dos avaliadores (média de K avaliadores) e o segundo tipo utiliza apenas a medida de um avaliador (BARTKO, 1966; FLEISS, JOSEPH L.; JACOB, 1973). De acordo com Koo (2016) essa seleção basicamente depende da maneira como o processo será conduzido, ou seja, se o objetivo final é utilizar o valor médio de avaliadores e suas avaliações como base de avaliação, deverá ser utilizada a “média de k avaliadores”. Por outro lado, se pretende usar o valor de um único avaliador como base na medição real, o tipo “avaliador único” deve ser o escolhido.

Já, a **Definição** considera se o interesse é calcular a concordância ou a confiabilidade. A concordância considera o interesse de verificar o coeficiente para diferentes avaliadores em um mesmo conjunto de dados que atribuíram a mesma nota para o mesmo recurso. A confiabilidade considera a pontuação fornecida pelos avaliadores para o mesmo conjunto de dados relacionando de maneira aditiva.

O presente trabalho utilizou o **Modelo** efeito misto bidirecional, pois os pares de usuários avaliadores são os únicos avaliadores de interesse, e existe o interesse de calcular o ICC para cada par de usuários avaliadores que possuem conjuntos de objetos comumente avaliados e que a escolha desses pares não é aleatória (e sim um consenso). A

escolha dos pares não foi aleatória porque no pré-processamento dos dados foi selecionado os pares de usuários de avaliadores que avaliaram os mesmos recursos, já a seleção aleatória não considera requisitos para a seleção da amostra. O **Tipo** utilizado foi o valor médio de K avaliadores, sendo assim, é possível utilizar o modelo de decisão ICC (3, K) que refere-se a fórmula do modelo efeito misto bidirecional, onde o 3 indica que o conjunto de dados analisados são classificados pelos mesmos avaliadores e o K indica que será considerado mais de um avaliador. E a **Definição** escolhida foi a concordância, pois o objetivo é calcular o coeficiente de concordância considerando a mensuração absoluta da mesma nota pelos pares de usuários avaliadores.

De acordo com Koo (2016) a fórmula computacional do ICC (3,K) utiliza as seguintes equações:

$$MS_B = \frac{k}{n-1} \sum_{j=1}^n (\bar{y}_{.j} - \bar{y}_{..})^2 \quad (3.1)$$

$$MS_E = \frac{k}{(n-1)(k-1)} \sum_{i=1}^k \sum_{j=1}^n (y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..})^2 \quad (3.2)$$

sendo que $\bar{y}_{.j}$ é a média das avaliações do avaliador j e $\bar{y}_{..}$ é a média de todas as avaliações. E $\bar{y}_{i.}$ é a média das avaliações de i documentos e y_{ij} é o valor da avaliação i do avaliador j.

Então, a fórmula computacional utilizada foi:

$$\frac{MS_B - MS_E}{MS_B} \quad (3.3)$$

a variável MS_E refere-se ao quadrado médio do erro e MS_B ao quadrado médio entre os sujeitos e o K é a quantidade de avaliadores dividida pelas medições.

Com o objetivo de evitar diferentes análises dos resultados, utiliza-se a interpretação de Fleiss, Levin e Paik (2003) a qual sugere que:

- Valores de ICC menores que 0,4 são indicativos de uma concordância baixa ou pobre.
- Valores de ICC entre 0,4 e 0,75 indicam uma concordância moderada ou satisfatória.
- Valores de ICC maiores ou iguais a 0,75 indicam uma concordância excelente entre avaliadores.

3.6 ANÁLISE DE CORRELAÇÃO E MEDIDAS DE ASSOCIAÇÃO

Muitas vezes é necessário entender o quanto uma variável possui relação com a outra para identificar o grau de significância entre elas (BARBETTA, 1994). Duas variáveis se associam quando são identificadas semelhanças na distribuição das avaliações (PARANHOS et al., 2014). Para isso, existem diversos métodos, seja para variáveis que seguem uma distribuição normal ou uma distribuição não conhecida. O método mais conhecido é o coeficiente de correlação Pearson, o coeficiente de Spearman e o coeficiente de Kendall .

O Coeficiente de Pearson mede a força e direção da associação linear entre duas variáveis relacionadas (HAUKE; KOSSOWSKI, 2011), podendo assumir a mesma interpretação apresentada na Figura 7.

O coeficiente de correlação de Kendall verifica o nível de associação entre duas variáveis (CHOK, 2010). De acordo com Kendall (1942), este é um coeficiente de correlação que pode ser utilizado como uma segunda alternativa para dados de classificações. O coeficiente é determinado por uma função matemática simples, sendo necessário obter o número de inversões de pares de objetos utilizados para transformar um posto de ordem para o outro (ABDI, 1955). Hauke e Kossowski (2011) afirmam que a principal vantagem de utilizar esse método é o fato de a distribuição ter características estatísticas mais significativas e de possuir uma interpretação que observa pontos concordantes e discordantes.

O coeficiente de correlação de Spearman é utilizado quando os dados apresentam uma distribuição muito assimétrica, é necessário utilizar uma abordagem não-paramétrica para calcular a correlação, o coeficiente mais utilizado é o de Spearman (BARBETTA, 1994). O coeficiente de Spearman verifica até que ponto é possível considerar que existe alguma relação entre as variáveis que sejam significativas, sem fazer suposições sobre a distribuição de frequência das variáveis (MANAGEMENT; NAUKOWE, 2011). Canovas (2009) define o coeficiente de Spearman como "uma medida de associação linear que usa os intervalos, números de pedidos de cada grupo de sujeitos e compara esses intervalos". De acordo com (BARBETTA, 1994) para calcular o coeficiente de Spearman, aplica-se a seguinte fórmula:

$$r_s = 1 - \frac{6 \sum D^2}{n(n^2 - 1)} \quad (3.4)$$

onde D refere-se à diferença entre os postos das duas variáveis, $\sum D^2$ a soma dos quadrados dos valores de D e n o tamanho da amostra.

Para qualquer amostra de dados do coeficiente, o intervalo será de -1 a 1. Como mostra a Figura 7, quanto mais próximo de 1 estiver o valor de r_s mais forte é a correlação. Por outro lado, quando os resultados forem descendentes ou não houver correlação, o r_s ficará próximo de 0 (BARBETTA, 1994).



Figura 7 – Força da correlação do valor de r_s

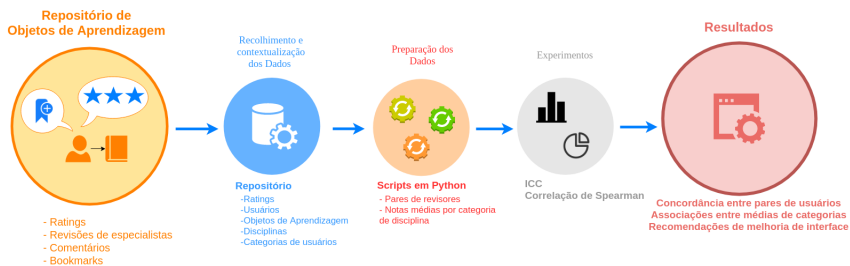
Para calcular a correlação foi escolhido o coeficiente de Spearman, pois os dados analisados apresentam uma distribuição muito assimétrica. Neste caso é necessário utilizar uma abordagem não-paramétrica.

4 PROCEDIMENTOS METODOLÓGICOS

Este capítulo apresenta a abordagem metodológica utilizada neste trabalho, descrevendo os processos utilizados em cada etapa de seu desenvolvimento.

A Figura 8 mostra o esboço geral da metodologia seguida.

Figura 8 – Esboço geral da metodologia



Fonte: Elaboração da autora.

Como mostra a Figura 8, o trabalho foi desenvolvido em cinco etapas. A primeira etapa foi realizado o estudo sobre o tema e a base de dados escolhida. Na segunda etapa foi efetuada a extração dos dados e a contextualização dos dados da base de dados Merlort. Foi realizada a descrição do conjunto de dados utilizados e os principais dados coletados para os estudos de casos. Na terceira etapa foi realizado o pré-processamento dos dados utilizando scripts em python, para gerar a tabela dos pares de usuários avaliadores e a tabela com as médias das notas das categorias. Na quarta etapa foi feito dois estudos de casos para analisar a concordância e a correlação, sendo que para cada análise foi necessário um pré-processamento específico de dados. Para isso foi feito o cálculo do coeficiente ICC e o Spearman. Por fim, foi feita a discussão dos resultados e realizada a recomendação de interface.

4.1 CONTEXTUALIZAÇÃO DOS DADOS

Os dados utilizados neste estudo foram extraídos do repositório Merlot (ano 2015), que possui um total de 15.886 avaliações de objetos de aprendizagem, sendo 11.823 avaliações realizadas por usuários e 4.064 avaliações realizadas por especialistas. Foi utilizada a base de

dados do ano de 2015, pois já havia sido extraído do repositório para utilizar os dados em outras pesquisas. Essa base de dados foi coletada usando um *crawler* da web que percorreu sistematicamente as páginas da web do repositório, sendo o mesmo utilizado no trabalho de Cechinel, Sánchez-Alonso e García-Barriocanal(2011).

O conjunto de dados possui um total de 28.098 usuários cadastrados. Sendo 12.020 professores universitários, 8.979 estudantes, os funcionários são formados por 2.240 usuários, 1071 são professores não universitários, 922 são administradores, os bibliotecários pertencem ao grupo de 470 usuários, 282 usuários são desenvolvedores, entre outros tipos de usuários. Existem 62.008 recursos cadastrados nas diferentes áreas do conhecimento: ciência e tecnologia, humanas, negócios, matemática, artes, educação. Sendo um total de 5.723 recursos avaliados pelos usuários por meio de notas que variam de 1 a 5.

4.2 EXTRAÇÃO DOS DADOS

Os dados da base Merlot foram coletados através de uma exportação de dados no formato SQL. Este arquivo SQL foi utilizado para popular uma base de dados local. Esta base de dados, que chamaremos de Dataset, possui 13 tabelas: `export`, `loauthors`, `locat_dat`, `locategories`, `locomments`, `lodata`, `lodata_has_primaryaud`, `loorganizations`, `loreviews`, `loval_com`, `lovaluation`, `loval_rev` e `users`. Na Figura 9 é apresentado o modelo relacional da base de dados.

cia entre os pares de usuários avaliadores. Conseqüentemente o recurso deve possuir mais de uma avaliação. O conjunto de dados apresentou um total de 1.452 avaliações. Também, foi verificado a quantidade de vezes que os usuários avaliaram os recursos do conjunto de dados. Foram excluídos todos os usuários que avaliaram menos que três recursos por não serem consideradas quantidades confiáveis para cálculos de concordância, pois podem ocasionar o acaso (KOO; LI, 2016) Foi agrupado, na mesma linha, cada recurso avaliado pelo mesmo usuário para não aparecer mais de uma vez o mesmo id. Ou seja, o usuário A, com id 123, que avaliou o recurso 1, recurso 2 e recurso 3 seria inserido na mesma linha da tabela.

Com esse conjunto foram selecionados os pares de usuários que avaliaram os mesmos recursos simultaneamente. Para identificar os pares de usuários avaliadores, foi realizada a comparação dos ids dos materiais avaliados. Se os ids dos materiais avaliados pelos pares de usuários avaliadores comparados fossem iguais, inseria as informações do segundo avaliador na mesma linha do primeiro avaliador. Com essa tabela formada, foi criado um novo id para identificar os trabalhos avaliados pelos mesmos pares. Com isso, foi possível agrupar todas as avaliações realizadas pelos mesmos pares de usuários avaliadores e salvá-las em um único arquivo, resultando em 179 pares de usuários. Na Tabela 2 é apresentado um exemplo do resultado obtido da tabela dos pares de usuários avaliadores.

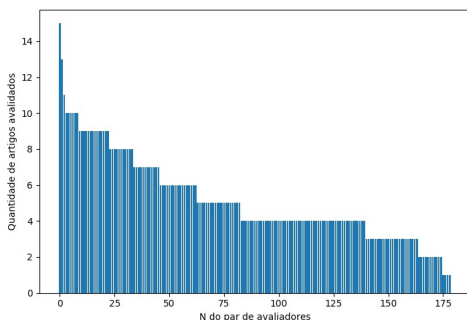
Tabela 2 – Exemplo da tabela dos pares de usuários avaliadores

Id_par	Material	Id_user_1	Rating_user_1	Id_user_2	Rating_user_2
18	Material A	10	4	12	3
18	Material C	10	5	12	4

A Figura 10 apresenta a distribuição da quantidade de recursos avaliados simultaneamente por pares (duplas) de avaliadores. Como é possível observar, apenas um par de usuários avaliou simultaneamente um total de 15 recursos idênticos (maior quantidade de recursos avaliados simultaneamente por uma mesma dupla de avaliadores). Desse total, 39 pares de usuários que avaliaram três ou menos recursos simultaneamente (≤ 3) foram excluídos do conjunto de dados por não serem consideradas quantidades confiáveis para cálculos de concordância (amostras pequenas de 3 ou menos recursos podem ocasionar o acaso (KOO; LI, 2016)). Ainda, 38 pares de usuários em que as avaliações consistiam somente de comentários (sem nota quantitativa) também foram excluídos, resultando em uma amostra de 102 pares de usuários ava-

liadores para o estudo realizado. Assim, foi possível identificar todos os recursos avaliados por cada par de usuários avaliadores. Sendo, 153 recursos distintos avaliados com 1.288 avaliações realizadas pelos 102 pares de usuários avaliadores.

Figura 10 – Quantidade de materiais avaliados por pares de usuários avaliadores



Fonte: Elaborada pela autora.

4.3.1 Pré-processamento para o estudo de caso I

Com os pares de usuários avaliadores separados com suas respectivas avaliações, o coeficiente ICC foi calculado para cada um dos pares e foram geradas estatísticas descritivas das concordâncias calculadas considerando, principalmente, três aspectos dos avaliadores: 1) disciplina de atuação preferencial, 2) subcategoria da disciplina de atuação preferencial e 3) classe de atuação profissional. Os algoritmos utilizados para gerar as tabelas mencionadas estão detalhados na Apêndice A.

1) Criar a tabela com a disciplina de atuação preferencial: o pré-processamento desta tabela também foi realizado através de *scripts* em Python. Para isso, foram identificadas duas das tabelas exportadas da base de dados que direcionavam para o nome da disciplina preferencial. A primeira tabela possui as informações completas dos usuários avaliadores. Então, foi realizada a comparação dos ids dos usuários. Se os ids fossem iguais era adicionado o id da disciplina preferencial do avaliador na tabela. A segunda tabela possui o id da disciplina preferencial e o nome da disciplina. Assim, foi possível comparar os ids da disciplina preferencial. Se os ids fossem iguais era adicionado o nome da disciplina na tabela. E teve como resultado

102 arquivos referentes aos pares de usuários avaliadores com o id do usuário, o nome dos materiais e a categoria da disciplina preferencial. O exemplo da tabela pode ser observado na Tabela 3

Tabela 3 – Exemplo da tabela de disciplina de atuação preferencial

Id_par	Material	Id_user_1	Rating_user_1	Cat_user_1	Id_user_2	Rating_user_2	Cat_user_2
18	Material A	10	4	Humanas	12	3	Tecnologia
18	Material B	10	5	Humanas	12	4	Tecnologia

2) Criar a tabela com a subcategoria da disciplina de atuação preferencial: para gerar a tabela da subcategoria da disciplina de atuação preferencial, foi realizado o pré-processamento pelo programa Knine ¹, pois o algoritmo desenvolvido em python não era otimizado o suficiente para processar os dados de forma eficiente. O Knine é uma solução de código aberto criada para descobrir informações nos dados, buscar novas informações ou prever resultados de forma eficiente. Foi utilizado o Knine para otimizar a execução do código. O Knine permite utilizar programação em blocos ou inserir scripts. Neste estudo foi utilizada a programação em blocos. Para iniciar o pré-processamento da tabela contendo a subcategoria da disciplina de atuação preferencial. Foram identificadas as tabelas necessárias para obter a subcategoria de cada avaliador. Após a identificação, as tabelas foram importadas para o programa. Por fim, foi exportada a tabela final contendo a subcategoria da disciplina preferencial de cada avaliador. O exemplo da tabela resultante pode ser observado na Tabela 4

Tabela 4 – Exemplo da tabela de subcategoria da disciplina de atuação preferencial

Id_par	Material	Id_user_1	Rating_user_1	Cat_user_1	SubCat_user_1	Id_user_2	Rating_user_2	SubCat_user_2	Cat_user_2
18	Material A	10	4	Humanas	Biologia	12	3	Educação	Ciências eTecnologia
18	Material B	10	5	Humanas	Biologia	12	4	Educação	Ciências eTecnologia

3) Criar a tabela com a classe de atuação profissional: o pré-processamento realizado para a classe de atuação profissional, foi similar ao descrito na disciplina preferencial. Também foi utilizado script em Python. Inicialmente foi identificada a tabela que possui o nome da classe de atuação profissional dos avaliadores. Em seguida, foi comparado o id dos usuários das duas tabelas. Quando os ids eram iguais, o nome da atuação profissional era inserido na tabela. O exemplo da tabela resultante está apresentado na Tabela 5

¹Mais informações em: <https://www.knime.com/knime-analytics-platform>

Tabela 5 – Exemplo da tabela de atuação profissional

Id_par	Material	Id_user_1	Rating_user_1	Atuacao_user_1	Id_user_2	Rating_user_2	Atuacao_user_2
18	Material A	10	4	Biólogo	12	3	Professor
18	Material B	10	5	Biólogo	12	4	Professor

4.3.2 Pré-processamento dos dados para o estudo de caso II

O segundo estudo de caso teve como objetivo verificar a associação entre as médias das avaliações dadas pelos avaliadores de cada uma das categorias de disciplina, permitindo a verificação da correlação entre as médias dos grupos. Para realizar o pré-processamento foi utilizada a biblioteca pandas do Python. Os algoritmos utilizados para o pré-processamento estão detalhados na Apêndice B. Inicialmente foi identificada a tabela que possui o nome e id do recurso avaliado, o id do usuário, a nota fornecida, o id e o nome da categoria. O exemplo a seguir mostra como foi organizada a tabela. A Tabela 6 mostra o conjunto de dados utilizados neste exemplo. O Material A foi avaliado três vezes por usuários de duas categorias diferentes. O Material B foi avaliado duas vezes por usuários das mesmas categorias. Já, o Material D foi avaliado duas vezes por usuários de categorias distintas. Então, o agrupamento dos dados apresentado na Tabela 7 juntaria todos os materiais iguais pertencentes às mesmas categorias.

Tabela 6 – Exemplo de dados utilizados

Recurso	Categoria do usuário	Nota
Material A	Categoria 1	4
Material A	Categoria 2	2
Material A	Categoria 1	3
Material B	Categoria 3	2
Material D	Categoria 4	1
Material B	Categoria 3	5
Material D	Categoria 3	4

O Material A avaliado por usuários da Categoria 1 ficou com Média 3,5, pois, as notas fornecidas para este material foram 4 e 3. Já, o Material A avaliado pelo usuário pertencente a Categoria 2 ficaria com média 2, pois, foi avaliado uma única vez pelos usuários desta categoria. O algoritmo utilizado está descrito no Apêndice B.

Tabela 7 – Exemplo do agrupamento dos dados

Recurso	Categoria do usuário	Média das notas
Material A	Categoria 1	3,5
Material A	Categoria 2	2
Material B	Categoria 3	3,5
Material D	Categoria 4	1
Material D	Categoria 3	4

Em seguida, foi organizada a tabela. A tabela resultante ficou formada pelas linhas representando as informações sobre as categorias dos recursos e outras 8 colunas com as categorias dos usuários que avaliaram os recursos e as médias das notas que os recursos obtiveram em cada uma das categorias, obtendo um total de 2.220 recursos avaliados. Por exemplo: se um recurso X pertencesse a categoria de tecnologia e tivesse recebido 3 notas, sendo o primeiro usuário de Negócios que forneceu nota 5, o segundo usuário de Computação que forneceu nota 4, e o terceiro usuário de Negócios dando nota 4. A tabela seria formada pelas categorias, a nota média dos avaliadores da área de Negócios que seria 4,5, a nota média dos avaliações da área de Matemática que seria 4, e as notas médias das outras áreas que nesse exemplo ficariam vazias.

4.4 CARREGAMENTO DOS DADOS

4.4.1 Carregamento dos dados para o Estudo de caso I

Com todas as tabelas criadas, foi possível calcular o grau de concordância entre os avaliadores de acordo com as variáveis selecionadas (disciplina preferencial, subcategoria da disciplina preferencial e classe de atuação profissional). Para a análise foram separados os coeficientes ICC em três categorias, pobre quando o valor do coeficiente ICC é menor que 0,4, satisfatória quando o coeficiente fica entre 0,4 e 0,75 e excelente quando os valores são maiores ou iguais a 0,75. Assim, foi possível gerar a tabela com os resultados finais contendo as porcentagens de concordância em cada categoria (pobre, satisfatória e excelente) e para cada variável selecionada (disciplina preferencial, subcategoria da disciplina preferencial e classe de atuação profissional). Ainda, foi identificado em cada caso quando os pares de usuários avaliadores pertenciam aos mesmo grupos e grupos diferentes.

Por fim, foi realizada a análise dos resultados apresentados na

Seção 5.1. Para testar se existe alguma associação entre as variáveis foi realizado o teste do qui-quadrado. De modo geral, buscou-se identificar as seguintes hipóteses do teste qui-quadrado H_0 as duas variáveis são independentes ou H_1 existe associação significativa entre elas (BARBETTA, 1994). Em todos os testes realizados, o nível de significância adotado foi de 5%. Para analisar o resultado do qui-quadrado foi realizada a comparação do p-value com o nível de significância.

4.4.2 Carregamento dos dados para o estudo de caso II

Com a tabela resultante do pré-processamento dos dados foram relacionadas as médias obtidas em cada categoria e aplicado um teste de correlação entre as colunas. Como as amostras não seguiram uma distribuição normal, uma análise não paramétrica foi realizada usando a correlação de Spearman para medir o grau de significância entre as médias das notas dadas pelos avaliadores e a categoria dos recursos, ou seja, se os avaliadores concordam ou não sobre a qualidade dos recursos. Com o resultado do coeficiente de Spearman foi possível montar uma matriz de correlação para gerar os gráficos de dispersão, além de calcular o p-value com objetivo de identificar se o r_s é significativo. Assim, foi possível gerar os gráficos de dispersão e realizar as análises dos resultados obtidos.

5 RESULTADOS

Nesta seção, apresentada-se a análise e discussão dos resultados obtidos no Estudo de caso I e II, descritos nas Seções 4.3.1 e 4.3.2, respectivamente.

5.1 ESTUDO DE CASO I - IDENTIFICAÇÃO DA CONCORDÂNCIA NAS AVALIAÇÕES DOS PARES DE USUÁRIOS AVALIADORES

O estudo de caso I busca analisar a concordância nas avaliações dos usuários de acordo com a disciplina de atuação preferencial, subcategoria da disciplina de atuação preferencial e atuação profissional. Para verificar se existe alguma concordância entre os usuários avaliadores da mesma área, considerou-se as situações em que ambos pertencem a uma área de atuação preferencial (mesma disciplina preferencial), além de verificar quando pertencem a áreas de atuação preferencial distintas. O mesmo teste foi realizado para verificar a existência de concordância considerando a atuação profissional e a subcategoria da disciplina de atuação preferencial.

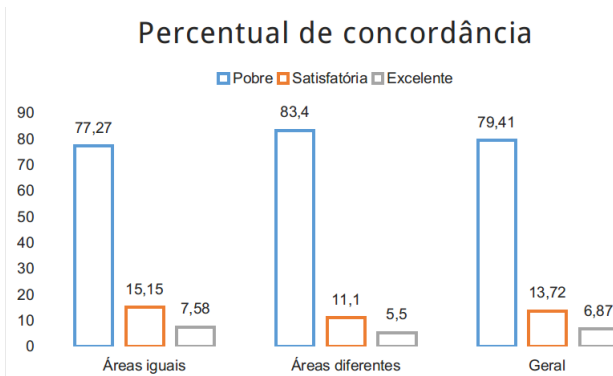
5.1.1 Concordância considerando as disciplinas de atuação preferencial

Usuários são registrados no Merlot com uma série de informações associadas, como a sua disciplina de atuação preferencial, sua atuação profissional, além de outros dados pessoais. As áreas de atuação preferencial são, de maneira geral, idênticas às categorias de disciplinas em que os objetos de aprendizagem são cadastrados, sendo elas: ciência e tecnologia, ciências sociais, humanas, negócios, matemática, artes e educação. Essas categorias preferencial, por sua vez, são subdivididas em subcategorias. Por exemplo, ciência e tecnologia é subdividida em Biologia, Física, Química, Tecnologia da Informação, entre outras. Nesse teste, consideramos somente a categoria de disciplina preferencial.

A Figura 11 apresenta os percentuais de concordância dos pares usuários de avaliadores considerando as situações em que ambos pertencem a uma área de atuação preferencial (mesma disciplina preferencial)

e quando pertencem a áreas de atuação preferencial distintas.

Figura 11 – Concordância geral por áreas preferencial iguais e diferentes



Fonte: Elaborada pela autora.

Como pode ser visto nesta figura, 79,41% dos pares de usuários avaliadores possuem uma concordância pobre ($ICC < 0,4$), enquanto 13,72% possuem concordância satisfatória ($0,4 \leq ICC < 0,7$) e apenas 6,87% possuem uma concordância excelente ($ICC \geq 0,7$). O percentual de concordâncias satisfatória e excelente entre os pares de usuários avaliadores que pertencem a uma mesma área de atuação preferencial é levemente maior do que os percentuais de concordância entre os pares de usuários avaliadores com áreas de atuação distintas. Enquanto no primeiro caso os percentuais de concordância satisfatória e excelente são de 15,15% e 7,58%, no segundo cenário estes percentuais caem para 11,1% e 5,5%.

Na Tabela 8 existe um determinado número em parênteses ao lado da porcentagem, esse número refere-se a quantidade de pares de usuários avaliadores que fazem parte dos grupos (pobre, satisfatória e excelente). Por exemplo, os avaliadores de áreas diferentes que pertencem ao grupo pobre equivale a 83,4% ou um total de 30 avaliadores. Com o teste do Qui-quadrado (χ^2) apresentado na Tabela 8 é possível concluir que tendencialmente ocorreu uma associação por variação casual, ou seja as variáveis não influenciam uma sobre a outra se existir alguma associação é considerado que ocorreu por acaso. Isto é confirmado pelos resultados obtidos e pelo teste Qui-quadrado (0,5233) que revelou não haver associação significativa ($p = 0,4694$) entre as avaliações dadas pelos pares de usuários avaliadores e a área de atuação,

mesmo sendo igual ou distinta.

Tabela 8 – Concordância geral por áreas preferencial iguais e diferentes

Tipo	Pobre			Satisfatória		Excelente		Total
	Va	x^2	Ve	Va	Va	x^2	Ve	
Avaliadores de áreas diferentes	83,4% (30)	0,07	28,59	11,1% (4)	5,5% (2)	0,27	7,41	36
Avaliadores de áreas iguais	77,27% (51)	0,04	52,41	15,15% (10)	7,58% (5)	0,15	13,59	66
Total de pares de usuários avaliadores	79,41% (81)	geral = 0,52		13,72% (14)	6,87% (7)	geral = 0,52		102

Legenda: Va =Valor absoluto; Ve = Valor esperado; x^2 = qui-quadrado; Valor em parênteses refere-se a quantidade de pares de usuários avaliadores

Vale mencionar que a qualidade da revisão por pares é criticada por muitos autores, pelo fato de identificar grandes divergências entre as avaliações fornecidas durante o processo de avaliação dos trabalhos científicos, levando a acreditar que o processo de revisão por pares não é totalmente eficiente (GRAYSON; MARY, 2002; KELLY; SADEGHIEH; ADELI, 2014). Assim, esta pesquisa contribui com o trabalho de (JENAL et al., 2012; SZKLO, 2006; JR; ROCHA; CHAMON, 2010) pelo fato de conseguir medir a concordância nas avaliações dos pares de usuários avaliadores. O teste do qui-quadrado foi realizado para a concordância *pobre*, *satisfatória* e *excelente*. Mas, o resultado do coeficiente pertencentes a categoria *satisfatória* e *excelente* foram somados com objetivo de discutir os coeficiente pobre ($< 0,4$) e os coeficientes com resultados considerados bons ($> 0,4$).

A Tabela 9 detalha os percentuais de concordância para pares de usuários avaliadores pertencentes a uma mesma área de atuação preferencial (disciplinas preferencial iguais). Apenas três áreas de atuação (ou disciplinas) possuem pares de usuários avaliadores de uma mesma área de atuação preferencial e que realizaram avaliações simultâneas, sendo elas: Humanas, Ciências e Tecnologia e Educação. A disciplina que possui um melhor grau de concordância entre os pares é a Ciências e Tecnologia com 15,2% de concordância satisfatória e 8,4% de concordância excelente. Embora essas associações ($p = 0,628821$) existam, elas não são muito fortes ($x^2 = 0,92$). Concluímos, então, que os dados não mostram evidência de associação ($p > 0,05$) significativas entre as variáveis analisadas. Comparando a Tabela 9 com a Tabela 8 é possível verificar percentuais parecidos, nota-se que os pares de usuários avaliadores discordam fortemente das avaliações fornecidas. Portanto, não é possível observar padrões de concordância nos dados apresentados. Se por um lado a pouca quantidade de pares nas áreas Humanas e Educação poderiam trazer resultados diferentes dos que foram apresentados na Tabela 8, com concordâncias satisfatórias e excelente, por

outro lado, este pequeno número também não auxilia nessa análise mais detalhada. Acredita-se que este pequeno número de indivíduos possam ter subáreas específicas diferentes entre si. Para tanto, faz-se uma avaliação das subáreas na Seção 5.1.2.

Tabela 9 – Concordância por áreas preferencial iguais

Áreas iguais	Pobre			Satisfatória	Excelente			Total
	Va	x^2	Ve	Va	Va	x^2	Ve	
Humanas	75% (3)	0,002	3,09	25% (1)		0,009	0,90	3
Ciências e Tecnologia	76,4% (45)	0,007	45,59	15,2% (9)	8,4% (5)	0,026	13,4	59
Educação	100% (3)	0,20	2,31			0,68	0,68	3
Total de pares de usuários avaliadores	77,27% (51)	geral = 0,92		15,15% (10)	7,58% (5)	geral = 0,92		66

Legenda: Va = Valor absoluto; Ve = Valor esperado; x^2 = qui-quadrado; Valor em parênteses refere-se a quantidade de pares de usuários avaliadores

A Tabela 10 apresenta o detalhamento dos percentuais de concordância para os pares de usuários avaliadores pertencentes a áreas de atuação de preferência distintas. Como já comentado, os percentuais de concordância estão concentrados em um ICC classificado como pobre. As exceções são as seguintes combinações das áreas de atuação: 1) Educação x Ciência e Tecnologia, 2) Negócios X Humanas, e 3) Negócios X Ciência e Tecnologia. O resultado do teste do qui-quadrado foi ($x^2 = 8,05$) e ($p = 0,70$). Como ($p > 0,05$) o teste de significância indicou que os pares de usuários avaliadores de áreas distintas não apresentaram nenhum padrão de concordância nas notas fornecidas. Isso pode ser confirmado na análise da Tabela 10, onde é possível observar que a porcentagem de pares de usuários avaliadores que concordam com as notas fornecidas é muito baixa, chegando a 16,6% ou 6 (quantidade de pares de usuários avaliadores apresentada em parênteses na tabela) pares de usuários avaliadores.

Tabela 10 – Concordância por áreas preferencial diferentes

Áreas diferentes	Pobre	Satisfatória	Excelente	Total
Humanas x Ciência e Tecnologia	100% (2)			2
Educação x Ciências Sociais	100% (2)			2
Educação x Ciência e Tecnologia	83,3% (10)	16,7% (2)		12
Ciência e Tecnologia x Ciências Sociais	100% (1)			1
Artes x Ciência e Tecnologia	100% (1)			1
Humanas x Ciências Sociais	100% (1)			1
Negócios x Humanas	50% (1)		50% (1)	2
Negócios x Ciência e Tecnologia	57,10% (4)	28,50 (2)	14,4% (1)	7
Educação x Serviços de apoio acadêmico	100% (2)			2
Educação x Humanas	100% (4)			4
Artes x Humanas	100% (1)			1
Matemática estatística x Ciências e Tecnologia	100% (1)			1
Total de pares de usuários avaliadores	83,4% (30)	11,1% (4)	5,5% (2)	36

Legenda: Valor em parênteses refere-se a quantidade de pares de usuários avaliadores

O conjunto de dados analisado apresentou uma grande discordância entre as avaliações dadas pelos pares de usuários avaliadores. Como mostram as Tabelas 8, 9 e 10. Essa discordância é levemente maior quando os usuários avaliadores pertencem a áreas de atuação preferencial distintas. Esse resultado já era esperado, entretanto, não esperávamos uma discordância tão grande entre os pares de usuários avaliadores de áreas de atuação iguais. Isso nos leva a um novo questionamento sobre as subcategorias dos avaliadores, que discutimos na Seção 5.1.2.

5.1.2 Concordância considerando as subcategorias das disciplinas de atuação preferencial

O usuário quando realiza o cadastro no Merlot tem a possibilidade de inserir a disciplina mais próxima ou a que ele mais se identifica em sua área de atuação profissional. Posteriormente, ele também pode incluir uma segunda disciplina (subcategoria ou subárea) que afunila a sua atuação. Por exemplo (exemplo retirado do merlor ¹), um avaliador da área preferencial de negócios pertencente ao tipo de usuário administrador, pode ter como primeira subcategoria (ou subárea) a disciplina de serviços de suporte acadêmico e como terceira disciplina o desenvolvimento de corpo docente, ou seja, as subcategorias definem certamente a atuação profissional dos usuários. Neste teste, consideramos apenas a subcategoria da disciplina preferencial.

¹Mais informações em <https://www.merlot.org/merlot/viewMember.htm?id=11397>

A Tabela 11 apresenta os percentuais de concordância dos pares de usuários avaliadores quando possuem a mesma subcategoria e a subcategoria distintas. Neste teste, uma amostra de 33,33% dos 102 pares de usuários avaliadores foram descartadas, pois os pares de usuários avaliadores não possuem subcategoria de atuação cadastrada ou apenas um dos avaliadores possui a subcategoria.

No geral, os resultados mostram uma maioria de discordância entre os avaliadores, assim como acontece com as áreas primárias dos avaliadores apresentados na Seção 5.1.1. No entanto, é possível visualizar na Tabela 11 uma maior quantidade de avaliadores que concordam de maneira satisfatória (30,44%) do que nos testes anteriores. Veja que, ainda, a maior porcentagem de pares de usuários avaliadores continua sendo a categoria pobre ($ICC < 0,4$) com 77,95%, enquanto a categoria excelente ($ICC \geq 0,7$) ainda pertence à menor quantidade de pares de usuários avaliadores com 7,35%. Ainda, os avaliadores de subáreas iguais apresentaram 8,69% de pares de usuários avaliadores na categoria excelente. Os avaliadores de áreas diferentes, relativamente discordam mais entre si, pois apenas 6,67% apresentaram concordância satisfatória e excelente.

Como a disciplina preferencial é a definição mais geral escolhida pelo usuário para determinar a área de atuação, pode ocorrer de usuários pertencerem a mesma categoria preferencial, mas a subcategoria diferente. Com os resultados apresentados na Seção 5.1.1 observa-se que a associação entre os avaliadores que possuem a mesma disciplina preferencial não apresentaram resultados satisfatórios. Mas, quando analisadas as subcategorias da disciplina preferencial percebe-se um aumento relativamente grande da porcentagem de pares de usuários avaliadores que concordam de maneira satisfatória e excelente. A amostra de pares de usuários avaliadores da mesma subcategoria e pertencente ao grupo satisfatório e excelente obteve um total de 39,13%. Diferentemente da categoria preferencial, a subcategoria restringi a área de atuação do usuário, ou seja, define melhor a área de atuação profissional. Por exemplo um avaliador pertencente a disciplina preferencial de educação pode ter como subcategoria a disciplina de professor. Então, provavelmente a análise das subáreas foi realizada por pares de usuários avaliadores que realmente pertencem a mesma área de atuação profissional. Por isso, a porcentagem de pares de usuários avaliadores concordando entre si nas subáreas iguais apresentaram uma associação maior. Por mais que a maioria dos pares de usuários avaliadores apresentaram uma concordância pobre, há quase 40% de concordância satisfatória e excelente. Resultado já esperado, pois acredita-se que

quanto mais parecida a área de atuação o nível de conhecimento sobre um determinado assunto será relativamente similar, levando a ter um nível de entendimento semelhante.

O teste do qui-quadrado foi aplicado da mesma forma que o teste anterior, ou seja, foi calculado para a concordância pobre e satisfatória mais excelente (foi somado os valores das duas colunas). Nestes resultados, o p - value ($p = 0,01521$) é menor que o nível de significância ($p < 0,05$), indicando que as variáveis possuem alguma associação. Em geral, quanto menor for o p, maior é a evidência da amostra. Observando a Tabela 11 percebe-se que a distribuição dos pares de usuários avaliadores das subáreas iguais e diferentes são semelhantes, com porcentagens muito próximas.

Tabela 11 – Concordância por subcategorias iguais e diferentes

Áreas iguais	Pobre			Satisfatória			Excelente			Total
	Va	x^2	Ve	Va	Va	x^2	Ve			
Avaliadores de áreas iguais	60,87% (14)	0,86	17,92	30,44% (7)	8,69% (2)	3,03	5,07	23		
Avaliadores de áreas diferentes	86,66% (39)	0,43	35,07	6,67% (3)	6,67% (3)	1,55	9,92	45		
Total de pares de usuários avaliadores	77,95% (53)	geral = 5,89		14,70% (10)	7,35% (5)	geral = 5,89		68		

Legenda: Va =Valor absoluto; Ve = Valor esperado; x^2 = qui-quadrado; Valor em parênteses refere-se a quantidade de pares de usuários avaliadores

A Tabela 12 mostra os resultados da concordância entre os pares de usuários avaliadores pertencentes às mesmas subcategorias de atuação. De modo geral, 60,9% da amostra possui uma concordância pobre. A subárea que apresentou maior concordância foi a de tecnologias da informação. Entre as 10 áreas de atuação apresentadas na tabela, 6 (psicologia, química, biologia, tecnologia da informação, física e ciências e tecnologia) apresentaram um ICC satisfatório ou excelente.

Com o resultado do teste qui-quadrado ($x^2 = 9,005$) e o $p = 0,43$ usando o nível de significância de 5%. Conclui-se, então, que as variáveis analisadas são independentes ($p > 0,05$), ou seja, não mostram evidências que existe alguma associação entre a subcategoria e as notas dadas pelos pares de usuários avaliadores.

Tabela 12 – Concordância por subcategorias iguais

Áreas iguais	Pobre			Satisfatória			Excelente			Total
	Va	x^2	Ve	Va	x^2	Ve	Va	x^2	Ve	
Psicologia		1,21	1,21	8,7% (2)			1,89		0,78	2
Química	8,7% (2)	0,01	1,82	4,3% (1)			0,02		1,17	3
Biologia	8,7% (2)	0,01	1,82	4,3% (1)			0,02		1,17	3
Inglês	4,3% (1)	0,02	0,60				0,39		0,39	1
Tecnologia da informação	13,3% (3)	0,11	3,65	8,7% (2)	4,3% (1)		0,18		2,34	6
Ciências da Computação	13,3% (3)	0,75	1,82				1,17		1,17	3
Professor	4,3% (1)	0,25	0,60				0,39		0,39	1
Física		0,60	0,60			4,3% (1)	0,94		0,39	1
Geociência	4,3% (1)	0,25	0,60				0,39		0,39	1
Ciências e Tecnologia	4,3% (1)	0,03	1,21	4,3% (1)			0,06		0,78	2
Total de pares usuários avaliadores	60,9% (14)	geral = 9,005		30,4% (7)	8,7% (2)	geral = 9,005				23

Legenda: Va = Valor absoluto; Ve = Valor esperado; x^2 = qui-quadrado; Valor em parênteses refere-se a quantidade de pares de usuários avaliadores

Como pode ser observado na Figura 13 a amostra de subáreas distintas apresenta uma discordância muito grande. Um total de 86,66% dos pares de usuários avaliadores possuem uma concordância pobre, indicando que os pares de usuários avaliadores interpretam de maneira distinta os recursos avaliados. Nestes resultados, o p ($p = 0,18665$) foi maior que o nível de significância ($p > 0,05$) definido para o teste. Conclui-se então, que não existe associação entre as notas dadas pelos pares de usuários avaliadores de subáreas distintas.

Tabela 13 – Concordância por subcategorias distintas

Áreas Diferentes	Pobre	Satisfatória	Excelente	Total
Inglês x Tecnologia da Informação	2,22% (1)			1
Inglês x História	4,44% (2)			2
Inglês x Geral	2,22% (1)			1
Inglês x Gestão	4,44% (2)		2,22% (1)	3
Inglês x Efetivo		4,44% (2)		2
Professor x Efetivo			2,22% (1)	1
Geral x Efetivo	2,22% (1)			1
Inglês x Música	2,22% (1)			1
Geral x Professor	4,44% (2)			2
Geral x Informações sobre o computador	2,22% (1)			1
Geociência x Geral	4,44% (2)			2
Geociência x Biologia	8,88% (4)			4
Geociência x Química	4,44% (2)			2
Química x Biologia	6,66% (3)			3
Ciências da Saúde x Física	4,44% (2)			2
Ciências da Saúde x Música	2,22% (1)			1
Biologia x Geral	2,22% (1)			1
Física x Química	4,44% (2)			2
Química x Geral	4,44% (2)		2,22% (1)	3
Tecnologia da Informação x Ciências e Tecnologia	6,66% (3)			3
Tecnologia da Informação x Informações sobre o computador	8,88% (4)	2,22% (1)		5
Professor x Serviços de Biblioteca e Informação	4,44% (2)			2
Total de pares de usuários avaliadores	86,66% (39)	6,67% (3)	6,67% (3)	45

Legenda: Valor em parênteses refere-se a quantidade de pares de usuários avaliadores

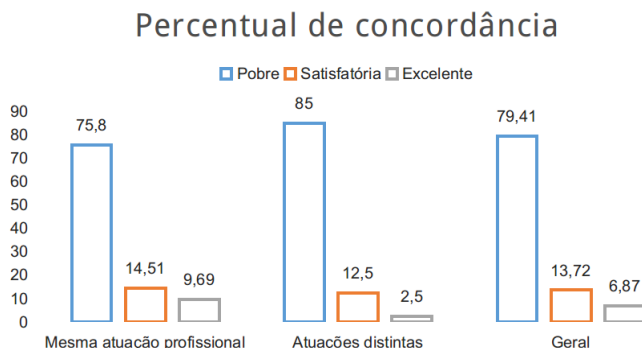
Os resultados deste segundo teste mostraram semelhanças com

o testes da área de atuação preferencial. Isto é, a amostra apresentou um grau de discordância muito grande e uma pequena parcela indicou equivalência nas avaliações. Além disso, a amostra analisada mostrou que as variáveis analisadas não apresentaram associações significativas.

5.1.3 Concordância considerando a atuação profissional

A terceira perspectiva de concordância avaliada é a relação da atuação profissional dos avaliadores. Exemplos de categorias de atuação profissional são: professor não universitário (*teacher*), estudante, profissional universitário (*faculty professional* - professor universitário, pesquisador), bibliotecário, consultor, e administrador. A Figura 12 apresenta os percentuais de concordância para avaliadores que possuem a mesma atuação profissional, atuações profissionais distintas, a concordância geral, a quantidade de pares de usuários avaliadores (número em parênteses) e o valor do qui-quadrado. Os resultados mostraram uma pobre concordância entre os pares de usuários avaliadores, independente de serem da mesma área de atuação profissional.

Figura 12 – Concordância geral considerando as áreas de atuação profissional



Fonte: Elaborada pela autora.

Com o objetivo de aprofundar a análise, a Tabela 14 apresenta o percentual e a quantidade entre parênteses para as três classes de atuação profissional com pares de usuários avaliadores que fizeram efetivamente alguma avaliação e são de uma mesma classe. Outras classes não formaram pares de usuários avaliadores para um mesmo OA. To-

dos os pares de usuários avaliadores professores não universitários e funcionários apresentaram uma concordância pobre ($ICC < 0,4$). Por outro lado, os avaliadores estudantes apresentaram um percentual de concordância maior, sendo que a soma dos percentuais de concordância satisfatório e excelente totaliza 28,31%, sendo esse grupo aquele que mais concorda nas avaliações realizadas. No entanto, o nível de significância ($p = 0,628821$) ainda é fraco ($p > 0,05$), o que reforça o resultado do Estudo de Caso I descrito na Seção 5.1 e os resultados na Seção 5. Mesmo analisando os pares de usuários avaliadores pela atuação profissional iguais e suas avaliações, ainda é possível observar que estes pares de usuários avaliadores possuem divergência nas avaliações fornecidas.

Tabela 14 – Concordância considerando a mesma área de atuação profissional

Mesma atuação profissional	Pobre			Satisfatória	Excelente			Total
	Va	x^2	Ve	Va	Va	x^2	Ve	
Professor não universitário	100% (8)	0,61	6,06					8
Estudante	71,69% (38)	0,11	40,1	16,98% (9)	11,33% (6)	0,36	12,8	53
Funcionário	100% (1)	0,07	0,75			0,24	0,24	1
Total de pares de usuários avaliadores	75,80% (47)	geral = 3,36		14,51% (9)	9,69% (6)	geral = 3,36		62

Legenda: Va = Valor absoluto; Ve = Valor esperado; x^2 = qui-quadrado; Valor em parênteses refere-se a quantidade de pares de usuários avaliadores

A Tabela 15, mostra a concordância entre avaliadores de áreas de atuação profissional distintas. Como pode ser visto, 82,5% dos pares apresentam uma concordância pobre entre as avaliações dadas. Com exceção de apenas 1 caso (Professor não universitário X Consultor), todos os demais casos em que existe alguma concordância satisfatória ou excelente, envolve a classe Estudante. A Tabela 15 mostra a mesma situação dos testes descritos anteriormente. A amostra apresenta um nível de significância ($p = 0,77$) onde ($p > 0,05$), demonstrando que a associação existente é muito baixa.

Tabela 15 – Concordância considerando áreas de atuação profissional distintas

Atuação profissional		Pobre			Satisfatória	Excelente			Total
		Va	x^2	Ve	Va	Va	x^2	Ve	
Professor não universitário	Profissional universitário	100% (6)	0,11	5,23			0,76	0,76	6
Estudante	Funcionário	87,5% (7)	0,00009	6,97	12,5% (1)		0,0006	1,02	8
Estudante	Professor não universitário	100% (3)	0,05	2,61			0,38	0,38	3
Estudante	Profissional universitário	77,78% (7)	0,33	8,71	22,22% (3)		2,30	1,2	10
Estudante	Administrador	75% (3)	0,06	3,48	25% (1)		0,46	0,51	4
Funcionário	Profissional universitário	100% (4)	0,07	3,48			0,51	0,51	4
Professor não universitário	Funcionário	100% (1)	0,01	0,87			0,12	0,12	1
Profissional universitário	Bibliotecário	100% (1)	0,01	0,87			0,12	0,12	1
Professor não universitário	Consultor	50% (1)	0,01	0,87		50% (1)	0,12	0,12	2
Bibliotecário	Profissional universitário	100% (1)	0,01	0,87			0,12	0,12	1
Total de pares de usuários avaliadores		85% (34)	geral = 5,67		12,5% (5)	2,5% (1)	geral = 5,67		40

Legenda: Va =Valor absoluto; Ve = Valor esperado; x^2 = qui-quadrado; Valor em parênteses refere-se a quantidade de pares de usuários avaliadores

De fato, não foi encontrada concordância significativa entre os pares de usuários avaliadores usando o cálculo do x^2 e o p-value apresentou-se sempre acima de 0,05. Deste modo, não se conseguiu definir o que cada avaliador analisa para entender quais critérios são utilizados nas avaliações realizadas.

5.2 ESTUDO DE CASO II - CORRELAÇÃO ENTRE AS MÉDIAS DAS NOTAS DAS CATEGORIAS

O estudo de caso II busca identificar a existência de correlação entre as médias das avaliações dadas pelos avaliadores de cada uma das categorias de disciplina, permitindo a verificação da correlação entre as médias dos grupos.

5.2.0.1 Correlação entre as médias das notas de cada categoria e as categorias dos usuários

Como o MERLOT permite que sejam separados os recursos e os avaliadores por categorias, foi coletada a amostra com a categoria principal dos mesmos. Com isso, foi possível identificar se existe alguma correlação entre as notas dadas, considerando a categoria de cada avaliador e recurso. Neste teste foi realizada uma análise não paramétrica usando a correlação de classificação de Spearman (r_s) para analisar se existe associação entre as avaliações. Na Figura 7 indica que a força de associação varia de -1 a 1. Quanto mais próximo de 0 o coeficiente de correlação estiver significa que não existe correlação e mais próximo de 1, indica que existe uma correlação significativa. A Tabela 16 apre-

senta as principais categorias (Educação, Humanas, Ciências Sociais, Ciências e Tecnologias, Serviços de apoio acadêmico, Negócios, Artes e Matemática e Estatística), o r_s e o p - value das médias das notas de cada categoria.

Como pode ser visto na Tabela 16 houve resultado do coeficiente $r_s = 1$, mas esse resultado somente foi obtido pelo fato de cruzar as mesmas categorias, por exemplo Educação x Educação ($r_s = 1$ e $p = 0$). São correlações inválidas pelo fato de comparar os mesmos dados. A categoria Ciências e Tecnologias foi a que apresentou os menores coeficientes de correlação, pois os coeficientes r_s foram os que ficaram mais próximo de zero e obteve um p muito grande, indicando que há uma grande discordância entre os avaliadores dessa categoria com os demais. No entanto, as categorias de artes x matemática e estatística ($r_s = 0,15$) e artes x humanas ($r_s = 0,09$) foram as categorias que apresentaram a maior correlação entre as avaliações dadas pelos avaliadores. Pelo fato do coeficiente de correlação ser o mais próximo de 1. Embora exista uma leve correlação elas não são significativas, ou seja, são correlações fracas, pois os coeficientes são pequenos.

Tabela 16 – Associação entre as notas e a categoria do usuário e recurso

	Média Educação		Média Humanas		Média Ciências Sociais		Média Ciências e Tecnologia	
	r_s	P	r_s	P	r_s	P	r_s	P
Média Educação	1	0	0,02	0,31	-0,04	0,03	-0,19	$7,76E^{-15}$
Média Humanas	0,02	0,31	1,0	0	0,03	0,16	-0,17	$7,76E^{-15}$
Média Ciências Sociais	-0,04	0,03	0,03	0,16	1,0	0	-0,16	$4,09E^{-13}$
Média Ciências e Tecnologias	-0,19	$7,76E^{-15}$	-0,17	$7,76E^{-15}$	-0,16	$4,09E^{-13}$	1,0	0
Média Serviços de apoio acadêmico	0,08	0,001	0,05	0,008	0,02	0,30	-0,07	0,0004
Média Negócios	-0,12	$1,05E^{-8}$	-0,02	0,19	0,001	0,94	0,06	0,005
Média Artes	0,01	0,39	0,09	$2,35E^{-13}$	0,06	0,005	-0,02	0,23
Média Matemática e Estatística	-0,009	0,68	-0,008	0,70	0,07	0,0005	-0,08	0,0002
	Média Serviços de apoio acadêmico		Média Negócios		Média Artes		Média Matemática e Estatística	
	r_s	P	r_s	P	r_s	P	r_s	P
Média Educação	0,08	0,0001	-0,12	$1,05E^{-8}$	0,01	0,39	-0,009	0,68
Média Humanas	0,05	0,008	-0,02	0,19	0,09	$2,35E^{-13}$	-0,008	0,70
Média Ciências Sociais	0,02	0,30	0,001	0,94	0,06	0,005	0,07	0,0005
Média Ciências e Tecnologias	-0,07	0,0004	0,06	0,005	-0,026	0,23	-0,08	0,0002
Média Serviços de apoio acadêmico	1,0	0	0,04	0,06	0,02	0,27	-0,02	0,20
Média Negócios	0,04	0,06	1,0	0	0,07	0,0004	0,04	0,02
Média Artes	0,02	0,27	0,07	0,0004	1,0	0	0,15	$4,76E^{-13}$
Média Matemática e Estatística	-0,02	0,20	0,04	0,02	0,15	$4,76E^{-13}$	1,0	0

Legenda: r_s = Coeficiente de Spearman; P = p - value

Uma maneira de visualizar os resultados apresentados na Tabela 16 e observar uma possível correlação entre as médias das notas fornecidas pelos pares de usuários avaliadores de cada categoria, é através do diagrama de dispersão. Como foi gerado um gráfico para cada relação de categorias, foram selecionados quatro gráficos de dispersão. As categorias que apresentaram os melhores resultados foram Artes x Humanas e Artes x Matemática e Estatística.

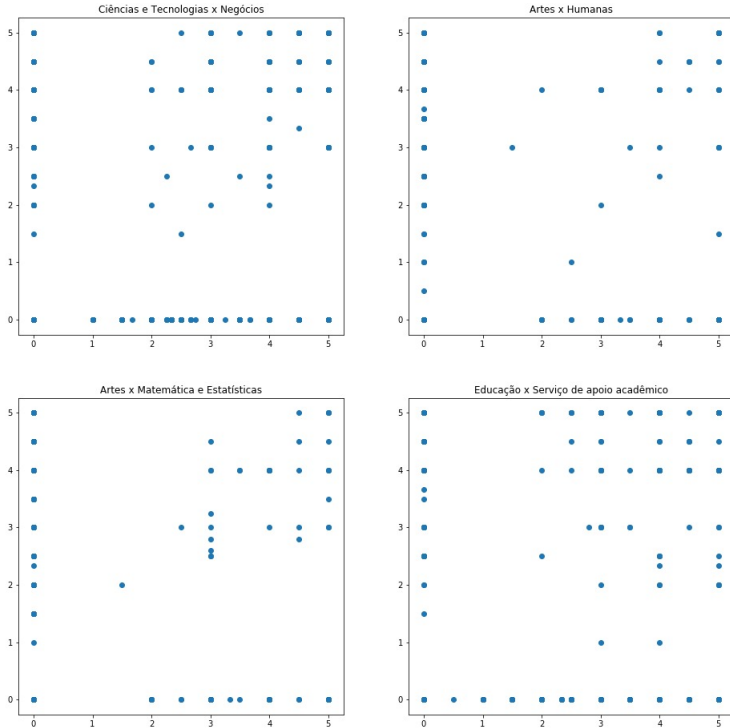
Como mostra a Figura 13, no diagrama de dispersão da média das notas da categoria Artes e Humanas existem diversos pontos

que formam uma reta na vertical indicando que as variáveis são independentes. O gráfico mostra uma leve dispersão dos pontos. Mas, a correlação observada é considerada fraca, ou seja, não existe influência das médias das notas à categoria dos avaliadores. Portanto, o coeficiente ($r_s = 0,09$) não é alto suficiente para considerar a existência de uma correlação significativa.

E o gráfico referente à categoria Artes e Matemática e Estatística apresentou uma correlação tão fraca que o gráfico não mostra nenhum sinal de aglomeração de pontos em linha diagonal. Além do coeficiente ($r_s = 0,15$) ser relativamente pequeno.

O diagrama de dispersão referente a disciplina de Ciências e Tecnologias e Negócios, e Educação e Serviços de Apoio Acadêmico também mostraram uma correlação fraca, mesmo demonstrando uma leve dispersão dos pontos o coeficiente de correlação é pequeno. O coeficiente obtido na correlação da disciplina de Ciências e Tecnologias e Negócios foi ($r_s = 0,06$) e na Educação e Serviços de Apoio Acadêmico foi ($r_s = 0,08$).

Figura 13 – Diagrama de dispersão da média das notas e as categorias



Fonte: Elaborada pela autora.

5.2.1 Recomendação da interface

Os ROA buscam sempre fornecer recursos de qualidade, mas também existe uma grande preocupação com a disponibilização ou organização das funcionalidades e informações (GARCÍA; ALONSO; MARÍN, 2014). Devido ao uso efetivo desses recursos acontecer pela coe-

rência da organização da interface (MESEGUER, 2001), muitos desenvolvedores incluem o estudo do design da informação como requisito no desenvolvimento do projeto, pois compreendem a necessidade de desenvolver sistemas que sejam úteis, utilizáveis e que consigam alcançar os usuários em potencial (BERGUES, 2013).

A visualização de informações nos últimos anos, aumentou significativamente em diversas áreas do conhecimento, sendo utilizada como estratégia no fornecimento de informações ao usuário (GARCÍA et al., 2018). O usuário consegue navegar e selecionar informações de interesse de maneira simples facilitando na absorção das informações fornecidas (MESEGUER, 2001). Meseguer (2001) afirma que a melhoria da interface do usuário deve ser realizada no contexto da qualidade do conteúdo, da atualização da informação e facilidade de uso. A representação visual do conhecimento é um exemplo de ferramenta que busca representar as informações de maneira simples e compacta por meio de gráficos que amenizam a carga cognitiva de atividades que necessitam de assimilação da informação, esse processo precisa de uma classificação estrutural para disponibilizar as informações com clareza independente da quantidade de informação (GARCIA et al., 2018).

De acordo com Islam e Bouwman (2016) deve ser considerado os seguintes ponto quando projetado um repositório:

- Desenvolvimento de uma estrutura de modelagem de informações que suporte a representação de conhecimento mais abrangente.
- O uso de representações padronizadas, para reduzir os esforços dos usuários na pesquisa e permitir a interoperabilidade com outros repositórios existente.
- Representações da informações que sejam interpretáveis facilmente.
- Desenvolvimento de taxonomias padronizada para fornecer consistência nos repositórios e facilitar no processo de indexação, navegação e busca.
- Desenvolvimento da interface do usuário simples, usual e eficaz na comunicação entre sistema e usuário.

Existem alguns estudos que buscam estratégias para transmitir as informações na interface do usuário de maneira simples e completas. Como por exemplo o trabalho de Marchionini (2006), Whote e Roth (2009) e García, Alonso (2014) que buscam identificar estratégias para melhorar o sistema de busca e navegação na interface. Já o estudo de

Martín-Moncunill (2013) identifica os repositórios que utilizam técnica de visualização para desenvolver a interface e avalia a interface do ponto de vista do usuário. Garcia et al. (2018) explica que o mal planejamento da interface do usuário pode levar ao impedimento de uso efetivo dos serviços fornecidos. Na sua pesquisa foi proposta uma estrutura capaz de avaliar e selecionar a interface do usuário mais adequada com o contexto do repositório.

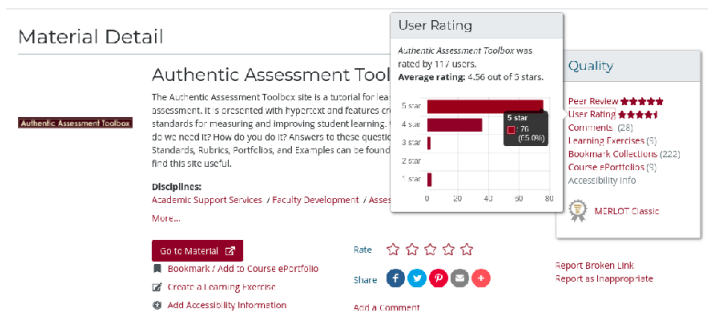
De acordo com Meseguer (2001) existem muitos sites que possuem conteúdo e funcionalidades de qualidade, mas a falta de padronização da interface e complexidade de uso leva ao sistema ser pouco intuitivo e o usuário enfrenta dificuldades para entender os esquemas de navegação e como as informações são organizadas. Portanto, a interface do usuário deve sofrer aprimoramento constante no design, pois as plataformas devem sempre verificar se está fornecendo informações completas, se as funcionalidades disponibilizadas funcionam corretamente e principalmente se os usuários estão utilizando o sistema, esse processo de aprimoramento é realizado com base em pesquisas e testes de usuários (NIELSEN, 1993). A presente pesquisa busca recomendar uma modificação na interface do usuário dos ROA, com o objetivo de apresentar informações mais completas sem aumentar a carga cognitiva de atividade ao usuários seguindo a recomendação de um modelo simples, fácil e utilizável.

Com a pesquisa realizada é possível identificar a importância das ferramentas que controlam a qualidade dos conteúdos disponibilizados nos ROA. O controle da qualidade em alguns ROA é garantida por comentários, avaliações de especialistas, classificações e entre outros. Além disso, é de suma importância garantir a qualidade dos serviços fornecidos e a apresentação das informações ao usuário final, com a finalidade de repassar de maneira coerente as informações solicitadas, distribuir as ferramentas de forma que facilite o uso da plataforma e apresentar os resultados com transparência e com amplo nível de interpretação por parte do usuário. O modelo de Delone e Mclean (2003) enfatiza que deve garantir a qualidade das informações, do sistema e dos serviços oferecidos para alcançar o sucesso nos ROA.

Como mostra a Figura 14 o Merlot fornece uma escala de 1 à 5 em formato de estrelas que determina a pontuação do material. Sendo separados pelos dois grupos de avaliadores (pares de usuários avaliadores especialistas e usuário avaliador). A quantidade de estrelas preenchidas referem-se a avaliação que o material obteve pelos usuários e especialistas, ou seja, se um material possuir 5 estrelas preenchidas ele possui uma popularidade alta e recebeu notas altas. Portanto, esse ma-

terial possui um conteúdo de qualidade. Essa pontuação apresentada na interface é uma nota geral que o material recebeu. Como o Merlot permite que diferentes tipos de usuário de distintas áreas de atuação forneçam pontuações a todos os materiais oferecidos sem nenhum requisito, pode ocorrer interpretações distintas da qualidade do material e pela nota geral apresentada ao usuário não é possível identificar tal informação.

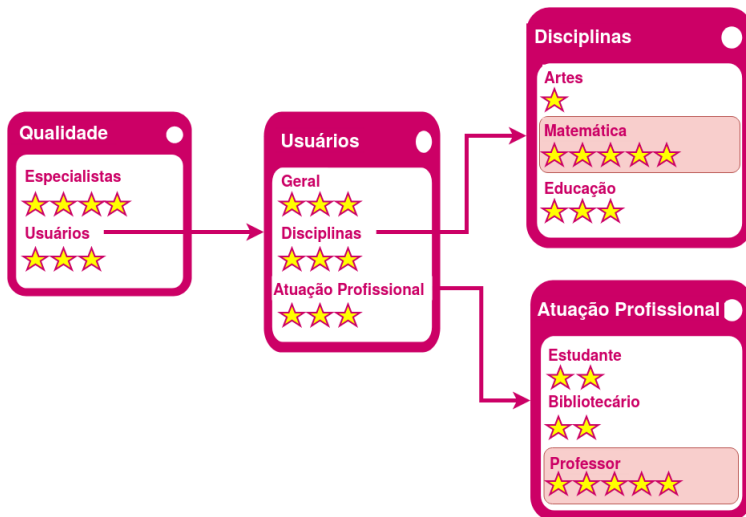
Figura 14 – Interface de avaliação da qualidade do Merlot



Fonte: Retirada do site Merlot.

Os resultados obtidos nos estudos de casos I e II mostraram uma grande discordância entre as notas dadas pelos pares de usuários avaliadores pela análise das categorias, a atuação profissional e médias das notas por categorias. Com esse resultado não fica clara a motivação que levou os usuários fornecerem as avaliações, indicando que os critérios utilizados são distintos. Mesmo que alguns resultados apresentaram uma leve concordância, por ser tão fraca não influenciou nas associações analisadas. Portanto, é importante apresentar essas informações de maneira clara para não induzir o usuário a confiar na qualidade do material apenas pela classificação geral. Com isso, foi elaborada um protótipo de como poderia ser apresentada a tela que contém as informações da qualidade dos ROA, Figura 15.

Figura 15 – Recomendação da interface



Fonte: Elaborada pela autora.

Já existem estudos que buscam apontar as melhores estratégias para transmitir as informações na interface dos ROA de maneira simples e completa, e aperfeiçoando os sistemas de busca e navegação nestas plataformas, uma vez que o mal planejamento da interface do usuário pode levar ao impedimento de uso efetivo dos serviços fornecidos (GAONA-GARCÍA et al., 2018).

Na Figura 14 mostra a tela principal do Merlot com as informações de qualidade do material. O usuário consegue navegar pela avaliação dos pares de usuários avaliadores, avaliação do usuário, comentários, recomendação, coleção de favoritos e cursos. A recomendação da interface apresentada, mostra melhorias na interface de apresentação da qualidade dos OA dentro dos ROA, de maneira a contemplar as discordâncias existentes dentro do processo avaliativo, Figura 15. Desta forma, a grande preocupação é a forma com que as informações e funcionalidades são disponibilizadas e organizadas na interface. Portanto, é importante apresentar a informação de maneira mais completa ao usuário para entender o tipo de avaliado que está fornecendo as notas ao material e qual a área de atuação profissional, assim o usuário consegue compreender a relação dos avaliadores com o material e aprofundar-se no entendimento da pontuação recebida.

A Figura 15 mostra uma nova forma de apresentação das in-

formações sobre a qualidade dos objetos de aprendizagem. O ROA deve, também, apresentar as informações segmentadas pelas diferentes categorias existentes para os avaliadores (e.g. área de preferência - disciplina, e atuação profissional, entre outras existentes). Com isso, as diferentes opiniões sobre a qualidade dos OA são apresentadas de forma mais transparente/detalhada para os usuários da plataforma, permitindo que os mesmos tenham acesso direto às opiniões de qualidade mais próximas de seu interesse ou perfil. Os repositórios poderiam, também, personalizar a apresentação da qualidade de um recurso para um determinado usuário, considerando especificamente as suas características em comparação com as características dos diferentes tipos de avaliadores e as pontuações que os mesmos deram aos recursos (resaltando na interface as pontuações mais similares com o seu perfil). No exemplo da Figura 15, o usuário que está logado na plataforma e realizando a consulta de um determinado OA é da área de Matemática e atua profissionalmente como Professor (categorias grifadas em rosa nas janelas).

O objetivo da recomendação da interface é fornecer informações mais completas para o usuário e entender como ocorre a avaliação por outros métodos, pois na pesquisa realizada a comunidade de usuários avaliadores demonstraram grande inconsistência nas avaliações do conjunto de dados avaliado. E com a média geral das avaliações os usuários não conseguem identificar que isso ocorre. Por isso, é importante existir mais de uma forma de avaliação dos recursos de uma plataforma, assim o usuário consegue comparar a pontuação recebida pelos demais grupos de avaliadores.

6 CONCLUSÃO

Na literatura foram encontradas pesquisas que utilizaram estratégias de verificação da concordância entre avaliadores, mas não com o mesmo objetivo da presente pesquisa.

A pesquisa utilizou dados recolhidos do repositório Merlot para medir a concordância e a correlação entre pares de usuários avaliadores que realizaram avaliações simultâneas sobre um mesmo conjunto de objetos de aprendizagem.

Com os resultados obtidos foi possível responder as questões de pesquisa levantada. **P1:** Qual a concordância geral entre pares de usuários avaliadores de objetos de aprendizagem? **R1:** Os resultados mostraram que os pares de usuários avaliadores possuem forte discordância com relação a qualidade dos objetos de aprendizagem que avaliaram simultaneamente. Dos 102 pares avaliados, 79,41% possuem uma concordância pobre, 13,72% possuem uma concordância satisfatória e apenas 6,87% possuem uma concordância excelente.

P2: Qual a concordância entre pares de usuários avaliadores considerando suas preferências por determinadas disciplinas, e subcategorias das disciplinas de atuação preferencial e suas áreas de atuação profissional? **R2:** As estatísticas descritivas dos ICCs calculados apontam que existe um leve aumento da concordância entre os pares de usuários avaliadores quando ambos pertencem a uma mesma área de atuação preferencial (mesma preferência por uma categoria de disciplina ou assunto curricular). Já, a subcategoria da disciplina de atuação preferencial (segunda disciplina selecionada para definir a atuação profissional) que apresentou maior concordância foi a tecnologia da informação e psicologia. Sobre a concordância considerando a área de atuação dos avaliadores, o grupo de avaliadores que mais apresenta concordância em suas avaliações é o de estudantes (com aproximadamente 28% dos pares com concordância satisfatória ou excelente). De maneira geral, o único grupo de avaliadores que concorda de maneira satisfatória ou excelente com outros grupos também é o grupo de avaliadores de estudantes (exceção para o caso específico dos pares formados pelos grupos professor não universitário e consultor).

P3: É possível identificar alguma associação significativa entre as notas dadas pelos avaliadores e as categorias dos avaliadores e dos recursos? **R3:** Os dados analisados apresentaram uma leve correlação, mas as correlações identificadas são muito fracas, indicando que não existe significância. As categorias (disciplinas) que apresentaram as

melhores correlações foram Artes x Humanas e Artes x Matemática e Estatística. Por mais que exista uma leve correlação as variáveis são consideradas independentes, ou seja, nenhuma das variáveis quando correlacionadas influenciam de maneira significativa os resultados.

Os resultados obtidos mostram uma grande discordância entre as notas dadas pelos pares de usuários avaliadores, e também pelos grupos de avaliadores pertencentes a diferentes categorias de disciplinas. Considerando que não existe um critério claro de avaliação para a comunidade de usuários no repositório estudado, não é possível precisar quais critérios estão sendo considerados nas avaliações realizadas, e é ainda mais difícil dizer qual a razão para os usuários fornecerem avaliações tão discordantes para um mesmo recurso. Isso ilustra a complexidade da análise de um processo avaliativo, de modo que afirmar se os resultados de uma avaliação são confiáveis ou não envolve muitos fatores.

A discordância encontrada levanta um problema que já vem sendo discutido que é a qualidade das revisões dos OA. É importante ressaltar que a qualidade da revisão por pares de especialistas já é criticada por diversos autores pela existência de grandes divergências sempre encontradas no processo de avaliação de trabalhos científicos (GRAYSON; MARY, 2002; KELLY; SADEGHIEH; ADELI, 2014). E na análise realizada nesta pesquisa das avaliações fornecidas pelos pares de usuários dos OA a discordância é muito forte, ressaltando que esse problema ocorre também nesse tipo de avaliação.

Com esses resultados, é importante levantar a questão dos processos de avaliações que existem hoje em dia se o nível de confiabilidade e concordância são confiáveis. Sem essa resposta não é possível identificar a eficácia desses processos e se as avaliações estão obtendo os resultados esperados.

Em busca de solucionar esse problema parcialmente, foi sugerido uma recomendação da interface distribuindo as notas conforme a área de estudo do usuário. Assim, o usuário consegue compreender melhor o perfil do usuário que forneceram as notas. Uma segunda solução seria a definição de critérios da qualidade para os usuários dos OA definir o quanto o material atende aos critérios definidos. Com essa solução os usuários teriam os mesmos critérios para avaliar determinado material.

Assim, a questão levantada pela pesquisa teve um resultado considerado ruim para o processo de qualidade dos materiais, pois mostrou que o processo de avaliação entre os usuários não é uma alternativa confiável atualmente.

Neste contexto, esta pesquisa corrobora os trabalhos de (JENAL

et al., 2012; SZKLO, 2006), pelo fato de conseguir detectar esta discordância entre as avaliações de pares de usuários, porém no contexto dos ROA.

Uma das limitações do presente trabalho é a pequena quantidade de pares de usuários avaliadores analisada e a consequente esparsidade das combinações de pares de usuários avaliadores. Em contrapartida, cabe ressaltar que o Merlot é um dos repositórios mais disseminados atualmente e que conta com uma das maiores quantidades de recursos, usuários e avaliações fornecidas pela sua comunidade. Nesse sentido, ainda que a quantidade de pares possa ser considerada pequena, ela é representativa de um dos maiores e mais importantes ROA atualmente existentes. Outra limitação é a possibilidade da discordância ser resultante da falta de critérios objetivos de qualidade a serem seguidos pela comunidade de usuário. Como trabalho futuro propõe-se conduzir experimentos futuros com uma amostra de usuários que avaliem os recursos utilizando as mesmas dimensões de qualidade utilizadas pelo Merlot para a sua avaliação por pares, sendo elas: 1) facilidade de uso, 2) qualidade de conteúdo, 3) potencial como ferramenta de ensino. Além da implementação desse tipo de recomendação da interface em ROAs e a avaliação de sua aceitação por parte da comunidade de usuários. E recomendar a estruturação da avaliação da qualidade por parte dos usuários em critérios específicos, ou seja, conseguir definir o motivo das notas fornecidas pelos usuários. Hoje em dia esse processo é feito de forma aberta, na qual não é possível identificar qual o critério que os usuários estão avaliando.

REFERÊNCIAS

- GARCÍA, P. A. G.; ALONSO, S. S.; MARÍN, C. E. M. Visualization of information: A proposal to improve the search and access to digital resources in repositories. *Ingenieria e Investigacion*, v. 34, n. 1, p. 83–89, 2014. ISSN 22488723.
- ABDI, H. The Kendall Rank Correlation Coefficient. 1955.
- ATENAS, J.; HAVEMANN, L. Questions of quality in repositories of open educational resources: a literature review. *Research in Learning Technology*, Association for Learning Technology, v. 22, Jul. 2014. <<https://journal.alt.ac.uk/index.php/rlt/article/view/1419>>.
- BARBETTA, P. A. *Estatística aplicada às ciências sociais*. 2. ed. Florianópolis: [s.n.], 1994. ISBN 8532800106.
- BARRIT, C.; ALDERMAN, F. L. Creating a reusable learning object strategy: leveraging information and learning in a knowledge economy. *Pfeiffer*, 2004.
- BARTKO, J. J. The intraclass correlation coefficient as a measure of reliability. *Psychological Reports*, SAGE Publications Sage CA: Los Angeles, CA, v. 19, n. 1, p. 3–11, 1966. <<https://doi.org/10.2466/pr0.1966.19.1.3>>.
- BENOS, D. J. et al. The ups and downs of peer review. In: *American Journal of Physiology - Advances in Physiology Education*. [S.l.: s.n.], 2007. v. 31, n. 2, p. 145–152. ISSN 15221229.
- BERGUES, L. P. C. y M. M. Usabilidad de los sitios Web , los métodos y las técnicas para la evaluación Usability of Web sites , methods and evaluation techniques. *Revista Cubana de Información en Ciencias de la Salud*, v. 24, n. 2, p. 176–194, 2013. ISSN 2307-2113.
- BLOCH, D. A.; KRAEMER, H. C. 2×2 kappa coefficients: Measures of agreement or association. *Biometrics*, [Wiley, International Biometric Society], v. 45, n. 1, p. 269–287, 1989. ISSN 0006341X, 15410420. <<http://www.jstor.org/stable/2532052>>.
- BURKE, R. H. Recommender systems: Survey and experiments. *User Model User-Adap Inter*, v. 12, p. 331–370, 2002.

CAFOLLA, R. Project merlot: Bringing peer review to web-based educational resources. In: WILLIS, D. A.; PRICE, J.; DAVIS, N. (Ed.). *Proceedings of Society for Information Technology & Teacher Education International Conference 2002*. Nashville, Tennessee, USA: Association for the Advancement of Computing in Education (AACE), 2002. p. 614–618. <<https://www.learntechlib.org/p/10579>>.

CARDINAELS, K. A dynamic learning object life cycle and its implications for automatic metadata generation. In: *Katholieke Universiteit Leuven – Faculteit Ingenieurswetenschappen A-bergenkasteel*. [s.n.], 2007. p. 163. ISBN 978-90-5682-805-9. <<https://lirias.kuleuven.be/handle/1979/882>>.

CECHINEL, C. *Objetos de Aprendizagem: introdução e fundamentos*. Editora UFABC, v. 1, p. 71–77, 2015.

CECHINEL, C. CIEB Estudos 5: “Modelos de curadoria de recursos educacionais digitais”. In: *Centro de Inovação Para A Educação Brasileira*. [s.n.], 2017. p. 48. <<http://www.cieb.net.br/cieb-estudos-5-modelos-de-curadoria-de-recursos-educacionais-digitais/>>.

CECHINEL, C.; SÁNCHEZ-ALONSO, S. Analyzing associations between the different ratings dimensions of the merlot repository. *Interdisciplinary Journal of E-Learning and Learning Objects*, Informing Science Institute, v. 7, n. 1, p. 1–9, 2011.

CECHINEL, C.; SÁNCHEZ-ALONSO, S.; GARCÍA-BARRIOCANAL, E. Statistical profiles of highly-rated learning objects. *Computers & Education*, Elsevier Science Ltd., GBR, v. 57, n. 1, p. 1255–1269, 2011. ISSN 0360-1315. <<https://doi.org/10.1016/j.compedu.2011.01.012>>.

CHARLESWORTH, A. et al. Development of Good Practice Guidelines for Repository Owners. p. 78, 2008.

CHOK, N. S. *Pearson’s versus Spearman’s and Krndall’s correlation coefficient for continuous data*. 4–5 p. Tese (Doutorado) — University of Pittsburgh, September 2010.

CLEMENTS, K. *Why Open Educational Resources Repositories Fail*. Jyväskylä studies in computing, 2016. 79 p. ISBN 9789513965884. <<http://urn.fi/URN:ISBN:978-951-39-6588-4>>.

CLEMENTS, K.; PAWLOWSKI, J.; MANOUSELIS, N. Open educational resources repositories literature review - Towards a

comprehensive quality approaches framework. *Computers in Human Behavior*, Elsevier Science Publishers B. V., v. 51, p. 1098–1106, 2015. ISSN 07475632. <<https://doi.org/10.1016/j.chb.2015.03.026>>.

CLEMENTS, K. I.; PAWLOWSKI, J. M. User-oriented quality for OER: Understanding teachers' views on re-use, quality, and trust. In: *Journal of Computer Assisted Learning*. [S.l.: s.n.], 2012. v. 28, n. 1, p. 4–14. ISSN 02664909.

COLLIS, B.; STRIJKER, A. Technology and human issues in reusing learning objects. *Journal of interactive media in education*, Ubiquity Press, v. 2004, n. 1, 2004. ISSN 1365-893X. Special Issue on the Educational Semantic Web.

DIMITRIADIS, Y. et al. New design approaches to repurposing open educational resources for collaborative learning using mediating artefacts. *Oroopenacuk*, p. 200–207, 2009. <<http://www.ascilite.org.au/conferences/auckland09>>.

DOWNES, S. Models for Sustainable Open Educational Resources. *Interdisciplinary Journal of Knowledge and Learning*, Informing Science Institute, Canadá, v. 3, n. 1, p. 29–44, 2007. ISSN 1552-2237. <<https://doi.org/10.28945/384>>.

EHLERS, U.-D.; PAWLOWSKI, J. M. *Handbook on Quality and Standardisation in E-Learning*. [S.l.]: Springer Science & Business Media, 2006. ISBN 9783540327875.

FLEISS, JOSEPH L., C.; JACOB. kappa Weighted Kappa Weighted kappa. p. 613–619, 1973.

GAONA-GARCÍA, P. A. et al. Usability of big data resources in visual search interfaces of repositories based on kos. In: *Proceedings of the 2018 2nd International Conference on Cloud and Big Data Computing*. New York, NY, USA: Association for Computing Machinery, 2018. (ICCBDC'18), p. 33–37. ISBN 9781450364744. <<https://doi.org/10.1145/3264560.3264567>>.

GARCIA, P. A. G. et al. Usability of big data resources in visual search interfaces of repositories based on kos. In: *Proceedings of the 2018 2nd International Conference on Cloud and Big Data Computing*. New York, NY, USA: Association for Computing Machinery, 2018. (ICCBDC'18), p. 33–37. ISBN 9781450364744. <<https://doi.org/10.1145/3264560.3264567>>.

GARCÍA, P. G. et al. Issues of visual search methods in digital repositories. In: *International Journal of Interactive Multimedia and Artificial Intelligence*. [S.l.: s.n.], 2018. v. 5, n. 3, p. 90. ISSN 1989-1660.

GARLICK, M. A review of creative commons and science commons. *EDUCAUSE Review*, v. 40, n. 5, p. 78–79, 2005. ISSN 1527-6619. <<https://www.learntechlib.org/p/99181>>.

GARZON, C. A. Tracing the creation and evaluation of accessible Open Educational Resources through learning analytics. In: *TDX (Tesis Doctorals en Xarxa)*. [s.n.], 2018. <<https://dugi-doc.udg.edu/handle/10256/15837?show=full>>.

GOLDBERG, D. et al. Using collaborative filtering to weave an information tapestry. *Commun. ACM*, Association for Computing Machinery, New York, NY, USA, v. 35, n. 12, p. 61–70, dez. 1992. ISSN 0001-0782. <<https://doi.org/10.1145/138859.138867>>.

GORDILLO, A.; BARRA, E.; QUEMADA, J. Towards a learning object pedagogical quality metric based on the lori evaluation model. In: *2014 IEEE Frontiers in Education Conference (FIE) Proceedings*. [S.l.: s.n.], 2014. p. 1–8. ISSN 2377-634X.

GRAHAM, M.; MILANOWSKI, A.; MILLER, J. *Measuring and Promoting Inter-Rater Agreement of Teacher and Principal Performance Inter-Rater Agreement of Teacher and Principal*. [S.l.: s.n.], 2012. ISBN 8882021513.

GRAYSON, L.; MARY, Q. Evidence based policy and the quality of evidence: Rethinking peer review. In: *ESRC UK Centre for Evidence Based Policy and Practice Lesley*. [s.n.], 2002. Working Paper 7. <<http://www.kcl.ac.uk/sspp/departments/politicaleconomy/research/cep/pubs/papers>>.

HAN, P. et al. Exposure and support of latent social networks among learning object repository users. In: *Journal of Universal Computer Science*. [S.l.: s.n.], 2008. v. 14, n. 10, p. 1717–1738.

HAUKE, J.; KOSSOWSKI, T. Comparison of values of pearson's and spearman's correlation coefficients on the same sets of data. *Quaestiones Geographicae*, v. 30, n. 2, p. 87–93, 2011. ISSN 0137477X.

HIGGS, P. E.; MEREDITH, S.; HAND, T. Technology for Sharing Flexible: Researching Learning Objects and Digital Rights

Management. *Australian Flexible Learning Framework Supporting*, v. 1, 2002.

HYLÉN, J. Open educational resources: Opportunities and challenges. In: *Proceedings of Open Education 2006: : Community, Culture, and Content*. Logan, UT, USA: [s.n.], 2006.

JENAL, S. et al. The peer review process : an integrative review of the literature *. In: *Acta Paulista de Enfermagem*. [S.l.: s.n.], 2012. v. 25, n. 5, p. 802–808. ISSN 0103-2100.

JR, H. V. M.; ROCHA, E. M.; CHAMON, W. Funcionamento e desempenho do sistema de Revisão por Pares. In: *Arq Bras Oftalmol*. [S.l.: s.n.], 2010. v. 73, n. 6, p. 487–488.

KELLY, J.; SADEGHIEH, T.; ADELI, K. Peer review in scientific publications : benefits , critiques , & a survival guide. In: *The Journal of the International Federation of Clinical CHEmistry and Laboratory Medicine - JIFCC*. [S.l.: s.n.], 2014. v. 25, n. 3, p. 227–243.

KOO, T. K.; LI, M. Y. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *Journal of chiropractic medicine*, Elsevier B.V., v. 15, n. 2, p. 155–163, 2016. ISSN 1556-3707. <<https://doi.org/10.1016/j.jcm.2016.02.012>>.

KRAEMER, H. C. Extension of the kappa coefficient. *Biometrics*, [Wiley, International Biometric Society], v. 36, n. 2, p. 207–216, 1980. ISSN 0006341X, 15410420. <<http://www.jstor.org/stable/2529972>>.

KUTLUCA, R. B. D. T. Review: Multimedia educational resource for learning and online teaching (merlot) (website review). *Turkish Online Journal of Distance Education*, Anadolu University, v. 11, n. 2, p. 235 – 237, 2010. ISSN 1302-6488.

LARSEN, K; LANCIN, S. The impact of ICT on tertiary education . In: OECD (Ed.). Washington DC: Michigan Conference “Advancing Knowledge and the Knowledge Economy, 2005.

LI, L. et al. Tutorial on use of intraclass correlation coefficients for assessing intertest reliability and its application in functional near-infrared spectroscopy-based brain imaging. *Journal of Biomedical Optics*, SPIE, v. 20, n. 5, p. 1 –13, 2015. ISSN 1083-3668. <<https://doi.org/10.1117/1.JBO.20.5.050801>>.

MACEDO, C. M. S. de. *Diretrizes para criação de objetos de aprendizagem acessíveis*. 76–99 p. Tese (Doutorado) — Universidade Federal de Santa Catarina (UFSC), 2010.

MACIEL, C. Educação a distância: ambientes virtuais de aprendizagem. In: *EduuFMNT Ambientes Virtuais de Aprendi*. [s.n.], 2018. p. 259. <<https://setec.ufmt.br/ri/handle/1/31>>.

MANAGEMENT, S.; NAUKOWE, B. W. Comparison of valus of pearson's and spearman's correlation coeficients on the same sets of data. *JOURNAL INFORMATION*, v. 30, n. 2, 2011.

MANOUSELIS, N.; SAMPSON, D. Recommendation of quality approaches for the european quality observatory. In: *IEEE International Conference on Advanced Learning Technologies, 2004. Proceedings*. [S.l.: s.n.], 2004. p. 1082–1083.

MATOS, D. A. S. Confiabilidade e concordância entre juízes: aplicações na área educacional. In: *Estudos em Avaliação Educacional*. São Paulo: [s.n.], 2014. v. 25, n. 59, p. 298–324. ISSN 1984-932X.

MESEGUER, A. P. Comunicación web: el valor de los contenidos de la Red. *Comunicar*, Revista Científica de Comunicación y Educación, n. 17, p. 93–96, 2001. ISSN 1134-3478. <<https://doi.org/10.3916/C17-2001-13>>.

MINGUILLÓN, J.; RODRÍGUEZ, M. E.; CONESA, J. Extending learning objects by means of social networking. In: LUO, X. et al. (Ed.). *Advances in Web-Based Learning – ICWL 2010*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2010. p. 220–229. ISBN 978-3-642-17407-0.

NESBIT, J. et al. A convergent participation model for evaluation of learning objects. *Canadian Journal of Learning and Technology / La revue canadienne de l'apprentissage et de la technologie*, Canadian Network for Innovation in Education, v. 28, n. 3, October 2002. ISSN 1499-6677. <<https://www.learntechlib.org/p/43018>>.

NESBIT, J. C.; LI, J. Web-Based Tools for Learning Object Evaluation. *Elearning*, v. 2, p. 3–8, 2004. <http://www.sfu.ca/jzli/publications/Nesbit_Li_2004.pdf>.

NEVEN, F.; DUVAL, E. Reusable learning objects: A survey of lom-based repositories. In: *Proceedings of the Tenth ACM International Conference on Multimedia*. New York, NY, USA: Association for

Computing Machinery, 2002. (MULTIMEDIA '02), p. 291–294. ISBN 158113620X. <<https://doi.org/10.1145/641007.641067>>.

NIELSEN, J. Iterative user-interface design. *Computer*, IEEE Computer Society Press, Washington, DC, USA, v. 26, n. 11, p. 32–41, nov. 1993. ISSN 0018-9162. <<https://doi.org/10.1109/2.241424>>.

OCHOA, X.; DUVAL, E. Towards automatic evaluation of learning object metadata quality. In: *Proceedings of the 2006 International Conference on Advances in Conceptual Modeling: Theory and Practice*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2006. (CoMoGIS'06), p. 372–381. ISBN 3540477039.

OCHOA, X.; DUVAL, E. Quantitative analysis of learning object repositories. *IEEE Transactions on Learning Technologies*, IEEE, v. 2, n. 3, p. 226–238, 2009. ISSN 19391382.

OLIVEIRA, E. R.; NELSON, M. A. V.; ISHITANI, L. Ciclo de Vida de Objetos de Aprendizagem baseado no Padrão SCORM. In: *SBIE - Simpósio Brasileiro de Informática na Educação*. [S.l.: s.n.], 2007. p. 217–226.

PARANHOS, R. et al. Desvendando os mistérios do coeficiente de correlação de Pearson: o retorno. *Revista Política Hoje*, Leviathan, São Paulo, n. 8, p. 66–95, 2014. ISSN 2237-4485.

PARK, J.-R.; TOSAKA, Y. Metadata quality control in digital repositories and collections: Criteria, semantics, and mechanisms. *Cataloging & Classification Quarterly*, Routledge, v. 48, n. 8, p. 696–715, 2010. <<https://doi.org/10.1080/01639374.2010.508711>>.

PAWLOWSKI, J. M. The quality adaptation model: Adaptation and adoption of the quality standard iso/iec 19796-1 for learning, education, and training. In: *Journal of Educational Technology & Society*. [S.l.: s.n.], 2007. v. 10, n. 2, p. 3–16. ISBN 11763647. ISSN 1436-4522.

PELAEZ, R. A. et al. A proposal of quality assessment of OER based on emergent technology. In: *2019 IEEE Global Engineering Education Conference (EDUCON)*. [S.l.: s.n.], 2019. p. 1114–1119. ISBN 9781538695067.

PERROCA, M. G.; GAIDZINSKI, R. R. Avaliando a confiabilidade interavaliadores de um instrumento para classificação de pacientes–coeficiente kappa. *Revista da Escola de Enfermagem da USP*, v. 37, n. 1, p. 72–80, 2003. ISSN 00806234.

POTTKER, L. *Arquitetura para recuperação de objetos de aprendizagem – uma abordagem baseada em agentes inteligentes e relevance feedback*. 53 p. Tese (Doutorado) — Universidade Estadual Paulista (UNESP), 2017.

SAMPSON, D. G.; ZERVAS, P. Learning object repositories as knowledge management systems. In: *Knowledge Management and E-Learning*. [S.l.: s.n.], 2013. v. 5, n. 2, p. 117–136. ISSN 20737904.

SHEA, P.; MCCALL, S.; OZDOGRU, A. Adoption of the multimedia educational resource for learning and online teaching (merlot) among higher education faculty: Evidence from the state university of new york learning network. In: *MERLOT Journal of Online Learning and Teaching*. [S.l.: s.n.], 2006. v. 2, n. 3.

SHOUKRI, M. M. *Measures of Interobserver Agreement and Reliability: Second Edition*. [S.l.]: CRC Press, 2010. 1–152 p. ISBN 9780429148590.

SHWETA; BAJPAI, R.; CHATURVEDI, H. K. Evaluation of inter-rater agreement and inter-rater reliability for observational data: An overview of concepts and methods. *Journal of the Indian Academy of Applied Psychology*, Indian Academy of Applied Psychology (IAAP), v. 41, n. 3, p. 20–27, 2015.

SZKLO, M. Quality of scientific articles. *Rev. Saúde Pública*, v. 40, p. 30–35, 2006.

TINSLEY, H. E. A.; WEISS, D. J. Interrater reliability and agreement. In: *Handbook of applied multivariate statistics and mathematical modeling*. San Diego, CA, US: Academic Press, 2000. p. 95–124.

TUOMI, I. Open educational resources and the transformation of education. In: *European Journal of Education*. [s.n.], 2013. v. 48, n. 1, p. 58–78.
<<https://onlinelibrary.wiley.com/doi/abs/10.1111/ejed.12019>>.

VARGO, J. et al. Learning object evaluation: Computer-mediated collaboration and inter-rater reliability. *International Journal of Computers and Applications*, Taylor & Francis, v. 25, n. 3, p. 198–205, 2003. <<https://doi.org/10.1080/1206212X.2003.11441703>>.

VET, H. *Observer Reliability and Agreement*. [S.l.]: American Cancer Society, 2014. ISBN 9781118445112.

VET, H. C. de et al. When to use agreement versus reliability measures. *Journal of Clinical Epidemiology*, Elsevier USA, v. 59, n. 10, p. 1033 – 1039, 2006. ISSN 0895-4356.

VIEIRA, A. L. e Ana Amaral e Julia Rocha e Marina Espindola e Roseli Cerny e D. Análise de critérios do processo de curadoria de red em plataformas digitais: uma perspectiva comparativa. In: *Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação - SBIE)*. [S.l.: s.n.], 2019. v. 30, n. 1, p. 21. ISSN 2316-6533.

WILEY, D.; GREEN, C.; SOARES, L. Dramatically Bringing down the Cost of Education with OER: How Open Education Resources Unlock the Door to Free Learning. In: *Center for American Progress*. [S.l.: s.n.], 2012. p. 1–5.

WILEY, D. A. Connecting learning objects to instructional design theory: A definition, a metaphor, and a taxonomy. v. 5, n. 1, p. 1–5, 2002. ISSN 00389145.

WILLIAM; FOUNDATION, F. H. White Paper: Open Educational Resources. n. November, p. 1–33, 2013. <[http://www.hewlett.org/sites/default/files/OER White Paper Nov 22 2013 Final_0.pdf](http://www.hewlett.org/sites/default/files/OER%20White%20Paper%20Nov%2022%202013%20Final_0.pdf)>.

WINDLE, R. J. et al. Sharing and reuse in OER: experiences gained from open reusable learning objects in health. In: *Journal of Interactive Media in Education*. [S.l.: s.n.], 2010. v. 2010. ISSN 1365-893X.

YUAN, L. et al. Open Educational Resources -opportunities and challenges for higher education Digital Commons Citation Open Educational Resources – Opportunities and Challenges for Higher Education. 2008. <http://digitalcommons.bolton.ac.uk/iec_reports/1>.

**APÊNDICE A - Algoritmos utilizados para o
pré-processamento das tabelas no estudo de caso I**

1) Algoritmo utilizado para inserir a disciplina de atuação preferencial.

```

1 import os
2 import pandas as pd
3
4 #Abriu a tabela que possui todas as informacoes dos usuarios
5 df1 = pd.read_csv('user_cat_dat_201903301617.csv', delimiter
6                 =',')
7 #Percorreu todos os pares de usuarios avaliadores
8 #Criou uma coluna para cada avaliador chamada "ID_cat-1"
9 for i in range(1,102):
10     name = 'clean-df/df'+str(i)+'.csv'
11     df = pd.read_csv(name, delimiter =',')
12     df['ID_cat-2'] = None
13     df['ID_cat-1'] = None
14     exists1 = 0
15     exists2 = 0
16 #Verifica se o id do usuario da tabela dos pares de usuarios
17 avaliadores sao iguais
18 #Se os ids forem iguais inseri o id da categoria do usuario
19 na coluna criada
20 for j in range(len(df1)):
21     if(df['users_ID_User_1'][0]== df1['ID_User'][[j]]):
22         if(exists1==0):
23             df['ID_cat-1'] = None
24             df['ID_cat-1'] = df1['ID_Cat'][[j]]
25             exists1 = 1
26         if(df['users_ID_User_2'][0] == df1['ID_User'][[j]]):
27             if(exists2 ==0):
28                 df['ID_cat-2'] = None
29                 df['ID_cat-2'] = df1['ID_Cat'][[j]]
30                 exists2=1
31             name2 = 'df-new-cate/df'+str(i)+'.csv'
32             df.to_csv(name2)

```

Algoritmo utilizado para inserir o nome da categoria na tabela dos pares de usuários avaliadores.

```

1 import os
2 import pandas as pd
3
4 #Abriu a tabela que possui o id e o nome da categoria
5 df1 = pd.read_csv('locategories_201903301617.csv', delimiter
6                 =',')
7 #Percorre todos os pares de usuarios avaliadores
8 #Verifica se o id da categoria inserido na tabela dos pares
9 de usuarios avaliadores e igual a outra tabela.
10 #Se forem iguais inseri o nome da categoria na coluna criada

```

```

9  for i in range(1,102):
10     name = 'df-new-cate/df'+ str(i)+'.csv'
11     df = pd.read_csv(name, delimiter=',')
12     for j in range(len(df1)):
13         if(df['ID_cat-1'][0]== df1['ID_Cat'][j]):
14             df['categories_user_1'] = None
15             df['categories_user_1'] = df1['Categorie'][j]
16         if(df['ID_cat-2'][0] == df1['ID_Cat'][j]):
17             df['categories_user_2'] = None
18             df['categories_user_2'] = df1['Categorie'][j]
19         name2 = 'df-new-cate-correta/df'+str(i)+'.csv'
20     df.to_csv(name2)

```

2) Algoritmo utilizado para inserir o tipo de membro

```

1  import os
2  import pandas as pd
3
4  #Abriu a tabela que possui as informacoes dos usuarios
5  df1 = pd.read_csv('users_201903301617.csv', delimiter=',')
6  #Percorreu as tabelas dos pares de usuarios avaliadores
7  #Se os ids fossem iguais verificava a coluna do tipo de
   membro e inseria o o tipo de membro referente ao id
   verificado
8  for i in range(1,102):
9     name = 'df-new-cate-correta/df'+ str(i)+'.csv'
10    df = pd.read_csv(name, delimiter=',')
11    for j in range(len(df1)):
12        if(df['users_ID_User_1'][0]== df1['ID_User'][j]):
13            df['memberType-1'] = None
14            df['memberType-1'] = df1['memberType'][j]
15        if(df['users_ID_User_2'][0] == df1['ID_User'][j]):
16            df['memberType-2'] = None
17            df['memberType-2'] = df1['memberType'][j]
18        name2 = 'df-with-member-type/df'+str(i)+'.csv'
19
20    df.to_csv(name2)

```

**APÊNDICE B – Algoritmos utilizados para o
pré-processamento das tabelas no estudo de caso II**

Algoritmo utilizado para agrupar os materiais e cada disciplina preferencial dos usuários avaliadores e calcular a média das notas fornecidas em cada disciplina preferencial.

```

1
2 import pandas as pd
3 #abriu o csv
4 df = pd.read_csv("tabela.csv", delimiter=",")
5 #criou um csv vazio
6 df2 = pd.DataFrame()
7 #agrupou o material e a categoria do usuario
8 df3 = pd.DataFrame({'count' : df.groupby( [ "Material", "
      name_cat", "Rating" ] ).size()).reset_index()
9 df2 = df.groupby(by=["Material", "name_cat"])["Rating"].mean
      ().reset_index()
10 #salvou cada dataframe em outra tabela
11 df2.to_csv("media_para_cada_recurso.csv")
12 df3.to_csv("agrupamento.csv")

```

Algoritmo utilizado para organizar a tabela com a média e a disciplina preferencial dos usuários.

```

1 #Abriu tabela com as medias das categorias
2 nv = pd.read_csv('agrupamento_media_autor_categoria.csv',
      delimiter=',')
3 #Criou um csv vazio
4 nova_tabela_organizada = pd.DataFrame()
5 #Criou uma nova coluna para salvar o nome do recurso
6 nova_tabela_organizada['NOME_DO_RECURSO'] =None
7 #Foi inserido o nome "Media_Autor_cat_Humanas" para
      identificar a media de cada categoria
8 for item in nv['Cat_autor']:
9     nova_tabela_organizada['Media_Autor_cat_'+item] = None
10 #Foi criada uma lista com todos os nomes criados
11 n=0
12 head_all = list(nova_tabela_organizada.columns[.])
13 # Percorreu a tabela que possui o nome do recurso, a media
      do recurso e a categoria do usuario.
14 # Para inserir todas as medias dos recursos em suas
      respectivas categorias
15 # Caso nao tenha avaliacao em uma determinada categoria foi
      adicionado NaN
16 for i in range(len(nv)):
17     if n == 0:
18         nova_tabela_organizada.loc[len(
          nova_tabela_organizada['NOME_DO_RECURSO']),
          'NOME_DO_RECURSO']= None
19         nova_tabela_organizada['NOME_DO_RECURSO'][i] = nv['
          Material'][i]

```

```

20     nova_tabela_organizada['Media_Autor_cat_'+nv[
        Cat_autor'][i]][i] = nv['Media_cat_autor'][i]
21 else:
22     if nv['Material'][i] == nova_tabela_organizada[
        NOME_DO_RECURSO'][len(nova_tabela_organizada[
        NOME_DO_RECURSO'])-1]:
23         nova_tabela_organizada['Media_Autor_cat_'+nv[
            Cat_autor'][i]][len(nova_tabela_organizada[
            NOME_DO_RECURSO'])-1] = nv['Media_cat_autor'
            ][i]
24     else:
25         nova_tabela_organizada.loc[len(
            nova_tabela_organizada[ 'NOME_DO_RECURSO' ]),
            NOME_DO_RECURSO'] = None
26         nova_tabela_organizada[ 'NOME_DO_RECURSO' ][len(
            nova_tabela_organizada[ 'NOME_DO_RECURSO' ])
            -1] = nv['Material'][i]
27         nova_tabela_organizada[ 'Media_Autor_cat_'+nv[
            Cat_autor'][i]][len(nova_tabela_organizada[
            NOME_DO_RECURSO' ]) -1] = nv['Media_cat_autor'
            ][i]
28     n += 1
29 # Criou um novo csv
30 df_final = pd.DataFrame(nova_tabela_organizada)
31 # Salvou a tabela
32 df_final.to_csv('tabela_medias_categorias.csv')

```

Algoritmo utilizado para criar a matriz de correlação de Spearman com as médias das notas de cada categoria.

```

1 import pandas as pd
2 import seaborn as sns
3 import matplotlib.pyplot as plt
4
5 #Abriu a tabela
6 df = pd.read_csv('tabela_medias_categorias.csv', delimiter="
        ,")
7 #Colocou 0 onde nao tinha nota
8 del df["Unnamed: 0"]
9 df.head()
10 df.fillna(0, inplace = True)
11 #Calculou o coeficiente de Spearman
12 corr = df.corr(method='spearman')
13 corr.to_csv("matrix.csv")
14 #Criando a matriz de correlacao
15 corr.style.background_gradient().set_precision(2)

```
