

Avaliação de Usabilidade de Sistemas de Gerenciamento de Referências Bibliográficas

Alternative Title: Usability Evaluation of Reference Management Systems

Ana Carolina T. Klock
PPGCA – CCT – UDESC
(Universidade do Estado de
Santa Catarina)
Joinville/SC, Brasil
actklock@gmail.com

Irmgard A. H. Campos
Nakazoni
PPGCA – CCT – UDESC
Joinville/SC, Brasil
irmgard.nakazoni@
hotmail.com

Isabela Gasparini
PPGCA, PPGECCMT –
CCT – UDESC
Joinville/SC, Brasil
isabela.gasparini@
udesc.br

Marcelo da S. Hounsell
PPGCA, PPGEEL – CCT
– UDESC
Joinville/SC, Brasil
marcelo.hounsell
@udesc.br

RESUMO

Devido ao grande volume de dados disponíveis, os pesquisadores encontram dificuldades para armazenar e organizar os trabalhos considerados relevantes para sua pesquisa. Os sistemas de gerenciamento de referências aperfeiçoam este processo, auxiliando na criação de um novo trabalho por meio da padronização e formatação adequada das citações e referências. O objetivo deste artigo é avaliar a usabilidade dos sistemas EndNote, Mendeley e Zotero. Foram utilizados métodos de avaliação sem e com a participação do usuário. Observou-se que os usuários do EndNote apresentaram a maior taxa de conclusão das tarefas solicitadas enquanto os usuários do Mendeley obtiveram a maior taxa de satisfação.

Palavras-Chave

Sistemas de Gerenciamento de Referências, Avaliação Heurística, Teste de Usabilidade, Questionário de Satisfação.

ABSTRACT

Due to the large volume of data available, researchers find it difficult to store and organize the relevant papers to their investigation. Reference management systems optimize this process, assisting in editing of a new paper through standardization and proper formatting of citations and references. The purpose of this paper is to evaluate the usability of EndNote, Mendeley and Zotero systems. Evaluation methods were used with and without user participation. It was observed that EndNote users had the highest completion rate of the requested tasks while Mendeley users achieved the highest satisfaction rate.

Categories and Subject Descriptors

H.5.2 [Information Interfaces and Presentation]: User Interfaces – *evaluation/methodology*.

General Terms

Measurement, Experimentation, Human Factors, Verification.

Keywords

References Management Systems/Software, Usability,

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

SBSI 2016, May 17–20, 2016, Florianópolis, Santa Catarina, Brazil.
Copyright SBC 2016.

Experiment, Heuristic Evaluation, Usability Testing.

1. INTRODUÇÃO

Atualmente, a utilização da internet viabiliza o acesso a uma gama de conhecimento advinda de diversas fontes, como as bases de periódicos. Logo, torna-se um desafio o gerenciamento das informações adquiridas de forma eficaz durante o estudo e pesquisa sobre um determinado tema [1].

Com o intuito de aperfeiçoar o gerenciamento destas informações, muitos pesquisadores recorrem a sistemas que auxiliam no processo de armazenamento e organização das fontes bibliográficas encontradas: os sistemas de gerenciamento de referências. Os gerenciadores de referências são sistemas que organizam os trabalhos selecionados de maneira automatizada, permitindo, entre outras funcionalidades, que os pesquisadores façam citações e referências no formato adequado.

Existem diversos sistemas disponíveis para o gerenciamento das referências bibliográficas, cada qual com suas características e formas de interação com o usuário. Assim sendo, o objetivo deste trabalho é avaliar a usabilidade dos gerenciadores de referências bibliográficas EndNote, Mendeley e Zotero por meio de diferentes métodos de avaliação da área de IHC (Interação Humano Computador). Essa avaliação visa primeiramente descobrir os problemas de usabilidade nas principais funcionalidades desses sistemas e também analisar o suporte dos métodos selecionados para a identificação dos problemas de usabilidade.

O trabalho está estruturado como segue: a seção 2 define o conceito de sistemas de gerenciamento de referências e descreve os sistemas escolhidos; a seção 3 apresenta alguns trabalhos relacionados; a seção 4 define os processos disponíveis para avaliar a usabilidade em um sistema; a seção 5 descreve o processo adotado para o desenvolvimento das avaliações e os resultados obtidos em cada etapa; a seção 6 realiza a discussão dos resultados; seguida pela seção 7 que conclui este trabalho.

2. GERENCIADORES DE REFERÊNCIAS

De acordo com Pampel & Dallmeier-Tiessen [2], um sistema gerenciador de referências é um conjunto de aplicativos que mantém uma base de dados de referências, contemplando as três etapas básicas de uma pesquisa: procurar, armazenar e escrever. Desta forma, o pesquisador pode coletar, organizar e utilizar estas referências em seus trabalhos. Além disso, as referências inseridas na base de dados podem ser reutilizadas [2]. Francese [3] corrobora com esta definição ao destacar duas funcionalidades primordiais dos gerenciadores de referências: 1) Construir uma base de dados de referências (para manter o controle e organizar a pesquisa) e 2) Auxiliar na formatação das citações e referências.

Neste artigo, foram selecionados três sistemas de gerenciamento de referências: EndNote, Mendeley e Zotero. Estes sistemas foram selecionados por sua popularidade no meio acadêmico, tanto em relação a sua utilização como enquanto objeto de estudo de diversas pesquisas. Nas próximas subseções os três sistemas são apresentados.

2.1 EndNote

O EndNote é um sistema de gerenciamento de referências criado em 1988 pela empresa Thompson Reuters, estando disponível em duas versões: a X7 e *Basic* [4]. A versão X7 apresenta custo de licença e está disponível para os sistemas operacionais Windows e Macintosh em versão *desktop*, podendo também ser acessada via *web* e aplicativo móvel (iOS) [4]. Com relação às funcionalidades, não apresenta limite de armazenamento de arquivos e referências, oferecendo mais de 6000 estilos de formatação bibliográfica, além da possibilidade de compartilhamento de conjuntos de referências (i.e., biblioteca) e integração com o Microsoft Word, entre outras funcionalidades [4]. A versão *Basic* é gratuita, entretanto apresenta limitações nas funcionalidades, podendo ser acessada somente via *web*. Esta versão contempla armazenamento de até 2GB de arquivos e 50000 referências, disponibilização de até 21 estilos de formatação bibliográfica, compartilhamento de referências e integração com o Microsoft Word [4].

2.2 Mendeley

O Mendeley foi criado em 2008 pela Elsevier e atualmente está disponível para os sistemas operacionais Windows, Linux e Macintosh em versão *desktop*, com possibilidade de acesso via *web* e aplicativo móvel (iOS e Android) [5]. Na criação de uma conta, os usuários têm direito a 2GB de armazenamento gratuito, sendo que este pode ser expandido para 5GB, 10GB ou espaço ilimitado através de respectivos pagamentos [5]. Com relação às funcionalidades, disponibiliza gratuitamente mais de 7500 estilos de formatação bibliográfica, apresentando integração com o Microsoft Word e LibreOffice, além de possibilitar o compartilhamento de biblioteca (de modo privado e público) [5].

2.3 Zotero

O Zotero foi criado pelo Roy Rosenzweig Center for History and New Media em 2006 e está disponível via *desktop* para os sistemas operacionais Windows, Linux e Macintosh, apresentando também acesso *web*, com extensão para o navegador Mozilla Firefox [6]. Os usuários têm direito a 300MB de armazenamento gratuito com opção de expansão para 2GB, 6GB e espaço ilimitado mediante pagamento [6]. Com relação às funcionalidades, disponibiliza mais de 6750 estilos de formatação bibliográfica, possui integração com o Microsoft Word, LibreOffice, OpenOffice e NeoOffice, permitindo também o compartilhamento de bibliotecas de forma privada e pública [6].

3. TRABALHOS RELACIONADOS

O trabalho de Yamakawa et al. [7] fez um estudo para comparar os mesmos três sistemas de gerenciamento de referências (EndNote, Mendeley e Zotero). Este trabalho evidenciou os principais benefícios e dificuldades de utilização em cada um, além de analisar suas funcionalidades por meio de um quadro comparativo. Como resultado, os autores concluíram que os três sistemas de gerenciamento de referências facilitam o trabalho dos pesquisadores por meio de ferramentas para buscas, organização e estrutura de artigos.

Por outro lado, Butros & Taylor [8] avaliaram as principais características dos sistemas EndNote, EndNote Web, RefWorks, Mendeley e Zotero, comparando os mesmos em relação as

vantagens e desvantagens, em termos de funcionalidades, de cada um. A proposta desta pesquisa foi auxiliar os pesquisadores na escolha e utilização das ferramentas avaliadas.

O trabalho de Klock et al. [9] analisou as funcionalidades sociais e colaborativas existentes nos sistemas de gerenciamento de referências bibliográficas EndNote, Mendeley e Zotero, comparando-os e identificando o mais abrangente. A ideia dos autores foi auxiliar os pesquisadores na escolha do sistema que mais atenda às suas necessidades.

O primeiro trabalho relacionado teve como objetivo evidenciar os principais benefícios e dificuldades de utilização em cada um dos gerenciadores, comparando suas funcionalidades. O segundo trabalho também avaliou as funcionalidades dos quatro gerenciadores de referência a fim de auxiliar na tomada de decisão na escolha dos usuários entre os mesmos. Já o terceiro trabalho buscou evidenciar as características sociais e de colaboração de três sistemas gerenciadores de referências. Entretanto, nenhum dos trabalhos encontrados avaliou os gerenciadores de referências bibliográficas EndNote, Mendeley e Zotero sob o ponto de vista da usabilidade.

4. AVALIAÇÃO DE USABILIDADE

A interface compreende os aspectos do sistema com os quais o usuário entra em contato de maneira física, perceptual ou conceitual [10]. O usuário utiliza a interface para interagir com o sistema, realizando tarefas no intuito de alcançar algum objetivo [11]. Assim, o alcance destes objetivos é condicionado à interface e à interação, que devem possuir características que assegurem alguns critérios de qualidade, como: usabilidade, experiência de usuário (UX), acessibilidade e comunicabilidade.

A usabilidade é um conjunto de componentes que garantem que os sistemas interativos sejam fáceis de aprender, efetivos no uso e agradáveis para seus usuários [12][13]. Existem diversas metas de usabilidade que auxiliam no processo de *design* de sistemas interativos, como: a facilidade de aprendizado, facilidade de recordação, facilidade de uso, eficiência, eficácia, segurança, utilidade e satisfação [12][13]. Dificilmente todas as metas conseguem ser alcançadas em sua totalidade e, por isso, cabe ao projetista da interface conhecer os usuários do sistema e suas necessidades para priorizar as metas mais relevantes [11].

A avaliação de usabilidade é uma parte do processo de *design*. Tal avaliação pode ser classificada conforme o ambiente onde é realizada, o envolvimento do usuário durante a avaliação e o nível de controle das atividades executadas [12]. Ela pode ser dividida em: avaliação sem a participação do usuário final (geralmente através da avaliação por inspeção e realizada por especialistas, ou pelas avaliações através de modelos preditivos), e avaliação com a participação do usuário (que pode ser dividida em avaliação em laboratório e em ambientes não controlados) [12].

A avaliação dos especialistas ocorre sem a participação dos usuários. Existem diversos subtipos dessa avaliação, como a avaliação heurística (onde os especialistas analisam se a interface está em conformidade com determinadas diretrizes) e o percurso cognitivo (que simula o ambiente e os problemas que o usuário final pode encontrar) [12]. A avaliação heurística é um método de inspeção de usabilidade realizado com base em uma série de diretrizes (denominadas heurísticas). Tais diretrizes auxiliam os avaliadores na inspeção sistemática da interface visando encontrar os problemas que afetem a usabilidade [13]. A avaliação heurística aloca de três a cinco especialistas (pois, diferentes especialistas tendem a encontrar diferentes tipos de problemas) que devem avaliar o processo interativo de forma independente.

Apesar de o processo avaliativo ser rápido, os especialistas podem não conhecer o contexto (e.g., tarefas, comportamento) dos usuários que irão utilizar o sistema [14]. Por isso, essa avaliação encontra problemas potenciais, que podem ser diferentes dos problemas reais enfrentados pelos usuários.

A avaliação com a participação do usuário em laboratório envolve alguns usuários utilizando o sistema em um ambiente controlado. Normalmente esta avaliação realiza os testes de usabilidade (onde os usuários são observados, tais como suas reações e seus erros, enquanto eles tentam realizar algumas tarefas no sistema) além da utilização de questionários (que auxiliam na caracterização do perfil dos usuários e na obtenção da opinião dos usuários) [12]. Os testes de usabilidade geralmente duram de uma a três horas e encontram os principais problemas referentes às atividades designadas, não abrangendo o sistema como um todo, nem os erros que seriam encontrados pelo seu uso contínuo [14]. O questionário é um método mais familiar e acessível para avaliar principalmente a satisfação do usuário, mas sua validade só pode ser conferida se realizado em conjunto com testes de usabilidade ou com avaliação por especialistas [14].

A avaliação com a participação do usuário em ambientes não controlados observa os usuários utilizando o sistema em seu cotidiano. Esse tipo de avaliação pretende visualizar como o produto seria utilizado no mundo real, tendo pouco ou nenhum controle sobre as atividades do usuário [12]. Uma das avaliações mais utilizadas neste caso é o estudo de campo, que tenta ser discreto e não afetar o cotidiano dos usuários [12]. Este tipo de avaliação costuma ser difícil de ser analisado e controlado, pois depende muito do *feedback* dos usuários ou precisa que o avaliador esteja presente quando algo importante acontecer durante os testes [12].

5. PROCESSO AVALIATIVO

A fim de analisar se os gerenciadores de referência escolhidos são fáceis de utilizar, eficazes e satisfatórios, foram realizadas avaliações sem a participação do usuário (avaliação heurística) e, posteriormente, com a participação do usuário (teste de usabilidade e questionários), conforme demonstra a Figura 1. Esse trabalho foi conduzido por nove estudantes da disciplina de IHC do Programa de Pós-Graduação em Computação Aplicada, da Universidade do Estado de Santa Catarina (UDESC) durante o primeiro semestre de 2015. Os estudantes/avaliadores foram separados randomicamente em três equipes, cada qual responsável por um dos sistemas avaliados: EndNote na versão X7.3.1, Mendeley na versão 1.13.2 e Zotero na versão 4.0.26.4.

Inicialmente, cada equipe realizou um levantamento das principais funcionalidades (i.e., atividades necessárias para que o usuário consiga armazenar, organizar, citar e referenciar trabalhos) que os sistemas tinham em comum. Apesar das diferenças entre terminologias, foram identificadas dezessete funcionalidades:

1. **Criar pasta:** permite criar um grupo (no EndNote), uma pasta (no Mendeley) ou uma coleção (no Zotero) para organizar os itens (e.g., artigos, livros) armazenados;
2. **Visualizar estrutura de pastas:** permite visualizar toda a estrutura das pastas criadas;
3. **Criar grupo:** permite criar uma pasta (no Mendeley) ou uma coleção (no Zotero) obrigatoriamente compartilhada com outros usuários. No EndNote, cada grupo (que equivale a uma pasta) pode ser compartilhado ou não;

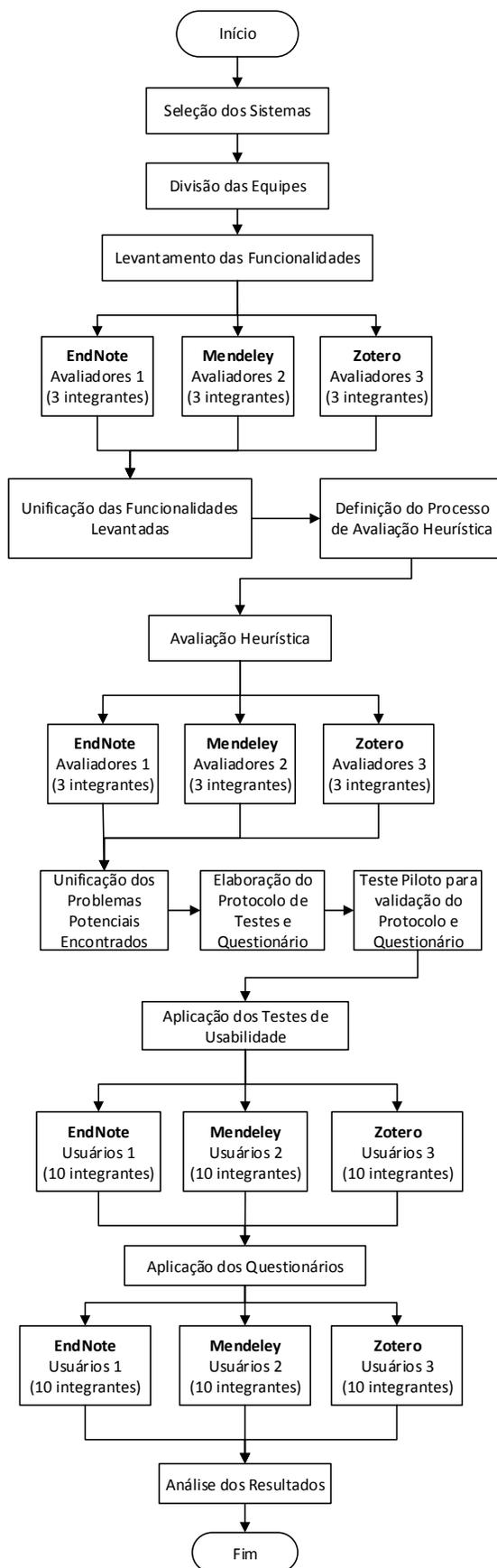


Figura 1. Etapas do Processo Avaliativo

4. **Importar arquivo bibliográfico:** permite importar um arquivo em um determinado formato bibliográfico (e.g., BibTeX, RIS) com dados de um ou mais trabalhos;
5. **Exportar arquivo bibliográfico:** permite exportar um ou mais trabalhos em arquivo em um determinado formato bibliográfico (e.g., BibTeX, RIS);
6. **Criar item:** permite criar um item para informar os dados de referência de um trabalho;
7. **Visualizar informações:** permite visualizar as informações sobre um item;
8. **Anexar arquivo do computador:** permite adicionar um arquivo (e.g., um arquivo no formato PDF) a um item (e.g., artigo de periódico, artigo de conferência);
9. **Visualizar arquivo anexado:** permite visualizar um arquivo que foi adicionado a um item;
10. **Adicionar marcadores:** permite adicionar marcadores (ou *tags*) a um item para facilitar futuras pesquisas e identificação melhor;
11. **Copiar citação ou referência na área de transferência:** permite exportar a citação ou a referência de um item conforme um estilo selecionado (no formato específico de uma conferência ou periódico) para colar em algum outro aplicativo;
12. **Realizar buscas simples:** permite buscar os itens armazenados com base em qualquer campo cadastrado ou nos marcadores relacionados ao item;
13. **Visualizar itens duplicados:** permite visualizar itens que foram cadastrados mais de uma vez;
14. **Adicionar citação ou referência em um documento do Microsoft Word:** permite adicionar citação ou referência em um determinado estilo diretamente em um documento criado no Microsoft Word, através de integração;
15. **Acessar conta:** permite acessar uma conta criada no *website* através do sistema, onde toda ação realizada (e.g., criação e organização de itens) é armazenada;
16. **Sincronizar conta manualmente:** permite buscar pastas, grupos e itens vinculados à uma conta e atualizar o sistema quando solicitado;
17. **Adicionar estilos para exportar citações e referências:** permite adicionar estilos específicos de conferências ou periódicos para que a citação e a referência sejam realizadas no formato correto.

Com base nas funcionalidades acima, a avaliação heurística foi realizada, conforme descreve a subseção a seguir.

5.1 Avaliação Heurística

A avaliação heurística é um método para avaliar a usabilidade de um sistema a partir de determinadas diretrizes, sendo uma alternativa rápida e de baixo custo para encontrar problemas na interface [15]. Um conjunto de diretrizes amplamente utilizado para realizar uma avaliação heurística são as dez heurísticas definidas por Nielsen [13]:

- **Visualização do estado do sistema:** O sistema deve manter o usuário informado sobre o que está acontecendo e fornecer um *feedback* apropriado conforme o tempo de espera;
- **Similaridade entre o sistema e o mundo real:** O sistema deve utilizar a linguagem do usuário, com palavras, frases e

conceitos familiares. Além disso, deve seguir as convenções existentes no mundo real, organizando a informação em uma ordem lógica e natural;

- **Controle do usuário e liberdade:** O sistema deve fornecer opções que permitam que o usuário abandone uma tarefa sem precisar passar por um diálogo muito extenso;
- **Padrões e consistência:** O sistema deve utilizar os mesmos termos para se referir a uma mesma coisa. Tais padrões também devem seguir convenções já existentes na plataforma que o usuário utiliza;
- **Prevenção de erros:** O sistema deve prever e evitar que um problema ocorra. Além de eliminar condições sujeitas à erros, o sistema deve confirmar se o usuário realmente deseja executar uma ação difícil ou impossível de ser desfeita;
- **Reconhecimento em vez de recordação:** O sistema deve minimizar a carga de memória do usuário. Os passos necessários para a utilização do sistema devem ser visíveis e de fácil reconhecimento, de forma que não obrigue a recordação de como o sistema funciona pelo usuário;
- **Flexibilidade e eficiência de uso:** O sistema deve fornecer atalhos, mesmo que invisíveis para os usuários mais novos, a fim de acelerar a iteração dos usuários mais experientes;
- **Estética e design minimalista:** O sistema deve exibir diálogos contendo apenas informações relevantes e necessárias;
- **Ajuda para que os usuários reconheçam, diagnostiquem e resolvam os erros:** O sistema deve fornecer mensagens de erro em uma linguagem simples (sem códigos), indicando precisamente o problema e sugerindo uma solução;
- **Ajuda e documentação:** O sistema deve disponibilizar ajuda e documentação para suprimir quaisquer dúvidas sobre a utilização. A documentação deve ser focada nas tarefas que o usuário realiza, indicar os passos necessários para concluir a tarefa, além de ser breve e fácil de procurar.

Os avaliadores, que geralmente são especialistas em IHC ou no método de avaliação, ou ainda estudiosos sobre as diretrizes a serem utilizadas [16], devem escolher as funcionalidades do sistema a serem inspecionadas para realizar a avaliação heurística. Na sequência, cada avaliador realiza a avaliação de forma isolada (i.e., sem contato com os demais avaliadores) e inspeciona cada funcionalidade pelo menos duas vezes (uma vez para entender o contexto geral e outra para inspecionar detalhes) [17]. Para cada problema encontrado, um grau de severidade entre zero e quatro é definido, sendo: 0 (não existe um problema de usabilidade), 1 (problema cosmético), 2 (problema pequeno), 3 (problema grande) e 4 (problema catastrófico) [17]. O grau de severidade serve para indicar a prioridade de correção do problema e considera-se os graus 3 e 4 como problemas que devem ser corrigidos antes da liberação do sistema. Ao final, todos os avaliadores apresentam seus resultados, discutem sobre os problemas encontrados por cada um e se os graus de severidade definidos estão adequados, unificando os resultados em um único relatório.

A avaliação heurística de cada sistema foi realizada em maio de 2015 pelos mesmos três estudantes que realizaram o levantamento das funcionalidades do sistema e, apesar de não serem considerados especialistas, possuíam um bom entendimento do procedimento e das heurísticas. As dezessete funcionalidades avaliadas foram descritas anteriormente, sendo que algumas das funcionalidades apresentaram mais de um problema ou apresentaram um problema que afetava várias heurísticas ao mesmo tempo. Por isto, a Tabela 1 identifica o

maior grau de severidade para o principal problema encontrado em cada funcionalidade, com base nos relatórios unificados dos sistemas.

As funcionalidades com mais problemas potenciais (i.e., com problemas nos três sistemas) foram as de criar pasta e de criar grupo. As funcionalidades com problemas mais graves (i.e., com graus de severidade igual a três ou quatro) foram a de criar grupo, de criar item e de citar ou referenciar no Microsoft Word.

Tabela 1. Severidade dos problemas em cada funcionalidade

Funcionalidade	EndNote	Mendeley	Zotero
Criar pasta	1	1	2
Visualizar estrutura	0	0	0
Criar grupo	3	2	2
Importar arquivo	0	0	0
Exportar arquivo	0	0	0
Criar item	3	0	2
Visualizar item	0	0	2
Anexar artigo	0	2	0
Visualizar artigo	0	2	1
Adicionar marcadores	2	0	2
Citar ou referenciar	0	2	0
Buscar item	2	0	0
Visualizar itens duplicados	0	0	2
Citar ou referenciar no Word	0	3	3
Acessar conta	0	0	1
Sincronizar conta	0	0	1
Adicionar estilo de citação	0	0	2

5.2 Teste de Usabilidade

Com base nos resultados da avaliação heurística, foi definido um protocolo unificado das funcionalidades a serem avaliadas em um teste de usabilidade com usuários potenciais deste tipo de sistema. Esse protocolo contemplou tanto funcionalidades críticas (i.e., com muitos problemas ou com problemas mais graves) quanto funcionalidades que precisavam ser executadas para conseguir realizar as funcionalidades críticas.

Durante um teste piloto, os avaliadores validaram o protocolo utilizando um dos sistemas que não o utilizado durante a avaliação heurística, anotando também o tempo para realizar cada tarefa. O protocolo final (adaptado a cada sistema devido à diferença entre terminologias), foi projetado para durar entre trinta minutos e uma hora, e foi composto das seguintes atividades:

1. Criar uma conta no *website* e acessá-la no sistema.
2. Criar um grupo chamado "OMITIDO".
3. Sincronizar a conta manualmente.
4. Criar uma pasta chamada "Metodologia de Pesquisa".
5. Criar um item dentro de "Metodologia da Pesquisa" com o título "Uma Reflexão sobre a Pesquisa em Ciência da Computação à Luz da Classificação das Ciências e do Método Científico", o autor "Raul Sidnei Wazlawick" e o ano "2010".
6. Anexar o arquivo "arquivo.pdf" disponível na área de trabalho do computador.
7. Visualizar o arquivo anexado.
8. Abrir o editor de texto Microsoft Word.
9. Citar o item criado com o estilo "*Chicago 16th (author-date)*".

10. Referenciar o item citado com o estilo "*Chicago 16th (author-date)*".

O artigo definido na atividade 5 foi escolhido devido a sua relevância e possível similaridade com o uso do sistema no cotidiano dos usuários potenciais que fariam o teste de usabilidade. O estilo definido na atividade 9 foi escolhido por ser um dos estilos instalados por padrão nos três sistemas. Os avaliadores também instalaram aplicativos para gravação da interação e das telas dos computadores em um laboratório da universidade e identificaram os computadores com um código para poder relacionar a gravação com o questionário, garantindo que os participantes fossem anônimos.

Os participantes deste teste de usabilidade foram alunos de graduação e de pós-graduação. Eles foram divididos em equipes, que tinha como requisito não ter conhecimento prévio do sistema. Após o agrupamento, os participantes foram solicitados a utilizar o sistema durante algumas semanas (utilizando seus computadores pessoais) e a explorar suas funcionalidades antes da realização do experimento. Somente os participantes que fizeram uso neste período e exploraram o sistema puderam fazer parte do experimento. Cada sistema foi utilizado em junho de 2015 por dez participantes, sendo quatro alunos de pós-graduação e seis alunos de graduação avaliando cada sistema.

Um termo de consentimento livre e esclarecido foi criado para explicar a finalidade da pesquisa, os procedimentos e a liberdade que os participantes tinham de cancelar sua participação quando desajassem. As orientações fornecidas explicaram que a avaliação era referente ao sistema e não sobre o conhecimento do participante e que, caso o participante não conseguisse realizar alguma atividade após algumas tentativas, poderia solicitar ajuda a um dos avaliadores, que os observavam e anotavam seus comportamentos e dificuldades. Esse termo foi assinado no início do teste de usabilidade e, na sequência, os participantes receberam o protocolo contendo orientações gerais e listando as tarefas a serem realizadas. Todo o experimento (teste de usabilidade e questionário de satisfação) com os participantes foi realizado nos laboratórios da universidade.

Tabela 2. Taxa de conclusão das tarefas solicitadas

Funcionalidade	EndNote	Mendeley	Zotero	Média
Acessar conta	100%	100%	100%	100%
Criar grupo	100%	100%	80%	93%
Sincronizar conta	80%	40%	70%	63%
Criar pasta	100%	100%	100%	100%
Criar item	100%	100%	100%	100%
Anexar artigo	100%	100%	90%	97%
Visualizar artigo	90%	90%	90%	90%
Abrir Microsoft Word	100%	100%	90%	97%
Citar item	70%	80%	60%	70%
Referenciar item	70%	80%	80%	77%
Média	91%	89%	86%	

A Tabela 2 demonstra o percentual de participantes que conseguiram concluir as atividades solicitadas no protocolo. Três participantes (um de cada sistema) realizaram o teste inteiramente na versão *web* em vez de acessar o sistema na área de trabalho. Nota-se que as tarefas com menor taxa de conclusão foram as de citação e referência de itens e a sincronização de conta. Apesar disso, o percentual mais baixo (40%) não se deve ao fato do participante não encontrar a opção (pois a maioria dos participantes direcionava o *mouse* e tentava clicar sobre a opção), mas sim devido ao sistema já realizar a sincronização automaticamente.

Para o teste de usabilidade, os dados de interação dos usuários com os sistemas foram gravados e armazenados em arquivos de vídeo, através do sistema CamStudio. Além disso, os avaliadores das etapas anteriores estavam presentes no teste, observando a interação do usuário e fazendo anotações de cada observação. Após a realização das tarefas propostas, os usuários responderam um questionário de satisfação.

Durante a análise da captura da tela, pode-se observar que alguns problemas ocorreram, mas não impediram que algumas tarefas fossem realizadas. No EndNote, por exemplo, observou-se que a maior dificuldade dos participantes estava em iniciar o uso do sistema, pois o sistema solicitava a criação de um arquivo para armazenar suas referências (não fica claro para o participante o objetivo de salvar o arquivo; isso não ocorre nos outros sistemas). No Mendeley, a maior dificuldade que os participantes tiveram foi ao criar um item, pois todos os participantes tentaram adicionar um novo item clicando com o botão direito sobre a pasta criada e não havia tal opção disponível. No Zotero, a maior dificuldade foi sincronizar o grupo manualmente, pois os participantes não encontravam a opção na interface com facilidade.

5.3 Questionário de Satisfação

Imediatamente após os testes de usabilidade, todos os participantes do teste de usabilidade foram convidados a responder um questionário de satisfação sobre o sistema. O questionário foi composto por perguntas que permitiram identificar suas características (e.g., faixa etária, curso), sua experiência anterior com sistemas para gerenciamento de referências, suas maiores dificuldades durante a utilização prévia ao experimento e durante o experimento, sua opinião sobre a interface e a interação do sistema, além dos pontos positivos e negativos sobre o sistema avaliado.

Em relação à faixa etária, quinze participantes tinham idades entre 18 e 25 anos, doze participantes tinham entre 26 e 35 anos e três participantes tinham entre 36 e 45 anos. Dos trinta participantes, onze já utilizavam algum sistema para gerenciamento de referências: dez utilizavam o Zotero e um utilizava o Mendeley. Vale lembrar que os sistemas avaliados por estes usuários não foram os mesmos que eles já utilizavam.

Entre as dificuldades encontradas durante exploração inicial do ambiente, antes da execução das tarefas do teste de usabilidade, os participantes do EndNote relataram que foi difícil encontrar os botões para as tarefas principais (reportado por quatro participantes), para realizar citação (reportado por dois participantes) e para importar arquivos bibliográficos (reportado por um participante). Os participantes do Mendeley relataram dificuldades para realizar citação (reportado por quatro participantes), criar item (reportado por dois participantes) e criar grupo (reportado por um participante). Os participantes do Zotero relataram a dificuldade para realizar citação (reportado por dois participantes), instalar (reportado por um participante) e descobrir como salvar um anexo ao item (reportado por um participante).

Já em relação às dificuldades encontradas durante o teste de usabilidade, com a execução das tarefas selecionadas, os participantes que utilizaram o EndNote relataram que tiveram dificuldade para criar grupos (reportado por três participantes), realizar citação (reportado por três participantes), acessar conta (reportado por dois participantes) e sincronizar conta (reportado por um participante), sendo que um dos participantes relatou que “no geral, as funções ficam em lugares de difícil acesso”. Os participantes que utilizaram o Mendeley relataram dificuldade para realizar citações e referências (seis participantes), criar grupos (reportado por três participantes), anexar arquivos

(reportado por dois participantes) e sincronizar a conta (reportado por um participante). Os participantes que utilizaram o Zotero relataram dificuldades para realizar citações e referências (reportado por quatro participantes) e sincronizar a conta (reportado por três participantes).

A Tabela 3 exibe os resultados obtidos pelo questionário em relação à usabilidade do sistema utilizado. Os aspectos foram avaliados por cada usuário com base em uma escala de Likert de 5 pontos: Discordo Totalmente (--), Discordo Parcialmente (-), Nem concordo nem discordo (-+), Concordo Parcialmente (+) e Concordo Totalmente (+). Os valores apresentados em cada coluna estão relacionadas a quantidade de participantes que respondeu cada escala.

Tabela 3. Quantidade de Respostas para cada Aspecto

		Aspecto Avaliado	--	-	-+	+	++
EndNote	É útil e tem as funções esperadas	0	1	1	4	4	
	É simples e intuitivo de usar	2	1	1	3	1	
	Atende as expectativas	0	2	2	4	2	
	Eu utilizaria novamente	3	1	3	1	2	
	Eu recomendaria para outros	2	1	2	3	2	
Mendeley	É útil e tem as funções esperadas	0	0	0	2	8	
	É simples e intuitivo de usar	0	0	2	4	4	
	Atende as expectativas	0	1	0	2	7	
	Eu utilizaria novamente	0	0	1	2	7	
Zotero	Eu recomendaria para outros	0	0	0	2	8	
	É útil e tem as funções esperadas	0	0	1	4	5	
	É simples e intuitivo de usar	0	3	2	5	0	
	Atende as expectativas	0	2	1	2	5	
	Eu utilizaria novamente	2	0	2	1	5	
Eu recomendaria para outros	2	0	1	2	5		

Os participantes que mais avaliaram positivamente os aspectos foram os que utilizaram o Mendeley (conforme pode ser visualizado, por exemplo, nos aspectos “É útil e tem as funcionalidades esperadas” e “Eu recomendaria para outros”, com 8 avaliações extremamente positivas), seguidos pelos participantes do Zotero e, por último, os participantes do EndNote.

6. DISCUSSÃO

Pode-se constatar que todos os sistemas possuem problemas, tanto na avaliação heurística, quanto no teste de usabilidade e no questionário de satisfação. O comparativo dos problemas encontrados durante os experimentos podem ser visualizados na Tabela 4, onde H equivale a “Avaliação Heurística”, U equivale a “Teste de Usabilidade” e S equivale a “Questionário de Satisfação”.

Tabela 4. Funcionalidades com algum problema identificado

Funcionalidade	EndNote			Mendeley			Zotero		
	H	U	S	H	U	S	H	U	S
Acessar conta			X				X		
Criar grupo	X		X	X		X	X	X	
Sincronizar conta		X	X		X	X	X	X	X
Criar pasta	X			X				X	
Criar item	X					X	X		
Anexar artigo				X		X		X	X
Visualizar artigo		X		X	X		X	X	
Citar item		X	X	X	X	X	X	X	X
Referenciar item		X	X	X	X	X	X	X	X

A funcionalidade mais crítica para todos os ambientes foi a criação de citações e de referências em um documento do Microsoft Word. Durante a avaliação heurística, este problema foi identificado como grande (grau de severidade 3) em dois dos três

sistemas e impediu que quase 30% dos participantes (no total) conseguissem realizar a funcionalidade durante o teste de usabilidade. Esse problema foi criticado de maneira geral em todos os sistemas durante o questionário de satisfação, sendo 5 críticas dos participantes que utilizaram o EndNote (dois na experiência prévia e três durante o experimento), 8 críticas dos participantes que utilizaram o Zotero (dois na experiência prévia e seis durante o experimento) e 10 críticas (i.e., a totalidade) dos participantes que utilizaram o Mendeley (quatro na experiência prévia e seis durante o experimento).

Entre as funcionalidades com problemas identificados, tanto sem quanto com a participação do usuário, pode-se citar a criação de grupo (identificados com problemas de grau 3, 2 e 2 e identificados por 3, 4 e 2 participantes para os sistemas EndNote, Mendeley e Zotero, respectivamente). No Zotero, também foi identificado o problema de sincronização de conta na avaliação heurística (classificado como problema cosmético), que também apareceu durante o teste de usabilidade (impedindo que 37% dos participantes conseguissem sincronizar) e no questionário (reportado por um participante). No Mendeley, foi identificado um problema ao anexar um artigo a um item na avaliação heurística (classificado como problema pequeno), sendo reportado por dois participantes no questionário de satisfação. No Mendeley e no Zotero foram identificados problemas na visualização de um artigo anexado (classificado como problema pequeno e cosmético, respectivamente) que impediu que um participante de cada sistema concluísse a tarefa no teste de usabilidade, além dos problemas para citar e referenciar itens.

Entre os problemas que foram identificados apenas durante a avaliação heurística, pode-se citar a criação de pasta e de item no EndNote (classificados como problema cosmético e grande, respectivamente), criar pasta no Mendeley (classificado como problema cosmético) e acessar conta, criar pasta e criar item no Zotero (classificados como problema cosmético e pequeno). Outros problemas identificados pelas avaliações com a participação dos usuários que não foram identificados durante a avaliação heurística no EndNote foram: o acesso a conta (reportado por dois participantes no questionário de satisfação), a sincronização da conta (reportado por um participante no questionário e não realizado por dois participantes nos testes de usabilidade) e a visualização do artigo (não realizado por um participante no teste de usabilidade), além da citação e referência de itens. No Mendeley, dois problemas foram identificados apenas durante testes com a participação do usuário: a sincronização da conta (não realizado por seis participantes no teste de usabilidade e reportado por um participante no questionário) e a criação de item (reportado por dois participantes no questionário). No Zotero, apenas um problema não foi identificado durante os testes sem a participação do usuário: anexar um arquivo a um item (não realizado por um participante durante o teste de usabilidade e reportado por três participantes no questionário).

A Figura 2 ilustra os problemas encontrados durante os experimentos realizados, classificados como: ambos (i.e., problemas que foram identificados tanto na avaliação sem quanto com a participação do usuário), sem participação (i.e., problemas que só foram encontrados na avaliação heurística) e com participação (problemas que foram encontrados durante o teste de usabilidade e/ou no questionário de satisfação). Os valores expostos no gráfico são distintos devido à diferença entre os problemas encontrados em cada sistema.

Para os sistemas Mendeley e Zotero, a maioria dos problemas encontrados durante a avaliação com a participação do usuário são, no geral, os mesmos identificados como problemas potenciais

na avaliação heurística. Entretanto, para o sistema EndNote, a maioria dos problemas foram encontrados apenas pelos participantes do experimento (i.e., teste de usabilidade e questionário).

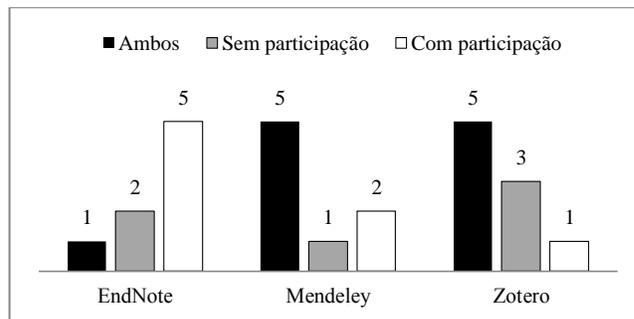


Figura 2. Problemas encontrados durante os experimentos

A facilidade de uso está relacionada com o esforço cognitivo para interagir com o sistema, e também com o número de erros cometidos durante esta interação [18]. Neste questionário focamos no primeiro aspecto. É possível mensurar a facilidade de uso através da afirmativa “É simples e intuitivo de usar” no questionário de satisfação. Como resultado, 40% dos participantes do EndNote, 80% dos participantes do Mendeley e 50% dos participantes do Zotero responderam positivamente (i.e., Concordo ou Concordo Totalmente), sendo que o Mendeley foi o único a não ter nenhuma resposta negativa (i.e., Discordo e Discordo Totalmente) para essa afirmativa. Dessa forma, pode-se inferir que Mendeley foi o sistema mais fácil de utilizar no contexto do teste realizado. Ao analisar a eficácia, entretanto, o EndNote foi o sistema com um maior percentual de conclusão das tarefas pelos participantes (91%), em comparação com o Mendeley (89%) e com o Zotero (86%).

Ao analisar a satisfação do usuário através das respostas das afirmativas do questionário “Eu utilizaria novamente” e “Eu recomendaria para outros”, o Mendeley foi o sistema com o maior percentual (95%) de respostas afirmativas (i.e., Concordo ou Concordo Totalmente), em comparação ao Zotero (65%) e ao EndNote (40%). Ainda, o Mendeley não obteve nenhuma resposta negativa (i.e., Discordo e Discordo Totalmente) para essas afirmativas, enquanto Zotero e EndNote tiveram 20% e 35% de respostas negativas, respectivamente.

Com isto, apesar do EndNote apresentar uma maior eficácia (pois teve o maior percentual de funcionalidades concluídas), o Mendeley teve o maior percentual de satisfação do usuário e também de facilidade de uso em relação aos outros sistemas. O Zotero, mesmo apresentando a menor eficácia, teve um maior percentual de satisfação e de facilidade de uso que o EndNote. Durante a escrita deste artigo, o sistema Zotero foi utilizado.

7. CONCLUSÃO

Neste trabalho foram realizadas avaliações de usabilidade sem e com a participação do usuário para comparar a usabilidade dos sistemas de gerenciamento de referências bibliográficas elencados, observando a eficácia, a facilidade de uso e a satisfação dos mesmos.

Durante a avaliação heurística, o sistema que apresentou menos problemas foi o EndNote, sendo também o que se mostrou mais eficaz às tarefas propostas. A maior parte dos problemas potenciais detectados durante a avaliação heurística também foram encontrados nos testes com a participação do usuário nos

sistemas Mendeley e Zotero, sendo que o mesmo não ocorreu para o EndNote.

Salienta-se que, apesar do EndNote ser o sistema com maior eficácia, o sistema que obteve a maior taxa de satisfação e de facilidade de uso foi o Mendeley.

Considerando os resultados obtidos, foi possível evidenciar a relevância e o caráter complementar da utilização conjunta das três ferramentas: avaliação heurística, teste de usabilidade e aplicação do questionário de avaliação, na construção de uma avaliação da usabilidade com maior acuracidade e confiabilidade.

Apesar disso, algumas limitações e ameaças à validade desse trabalho podem ser citadas, como o fato da avaliação heurística ter sido realizada com três avaliadores distintos em cada sistema (e não os mesmos três avaliadores para todos os sistemas), sendo que esses avaliadores não eram especialistas, mas sim alunos da disciplina de IHC. Os testes de usabilidade foram realizados somente nas tarefas selecionadas durante a avaliação heurística e com uma quantidade pequena de participantes (dez alunos para cada sistema) e ainda todos participantes do experimento eram da área da computação (i.e. tinham prática com tecnologia), o que pode influenciar na facilidade de aprendizagem e uso das ferramentas. Todos os participantes nunca tinham utilizado o sistema selecionado para seu uso, e foram solicitados a explorar a ferramenta antes do teste, porém, não foram realizados nenhum tipo de treinamento prévio (i.e. cada participante pôde utilizar o sistema anteriormente o quanto quisesse), deste modo não podemos assegurar que todos exploraram o sistema com a mesma profundidade e abrangência.

Como trabalho futuro, pretende-se também analisar os tempos que cada participante levou para realizar cada tarefa, podendo-se mensurar também a eficiência dos três sistemas. Além disso, outros públicos e tarefas serão explorados.

8. REFERÊNCIAS

- [1] Duong, K. (2010). Rolling Out Zotero Across Campus as a Part of a Science Librarian's Outreach Efforts. *Science & Technology Libraries*, v. 29, n. 4, p. 315–324.
- [2] Pampel, H. e Dallmeier-Tiessen, S. (2014). Open Research Data: From Vision to Practice. In: Bartling, S.; Friesike, S. [Eds.]. *Opening Science: The Evolving Guide on How the Internet is Changing Research, Collaboration and Scholarly Publishing*. Cham: Springer International Publishing. p. 213–224.
- [3] Francese, E. (2011). The Usage of Reference Management Software (RMS) in an Academic Environment: a Survey at Tallinn University. *Session on Divergence and Convergence: Information Work in Digital Cultural Memory Institutions*,
- [4] EndNote Official Website Basic <http://endnote.com/product-details/basic>, [accessed on Feb 12].
- [5] Mendeley Official Website - Overview about Mendeley (2016) <https://www.mendeley.com/features>, [accessed on Feb 12].
- [6] Zotero Official Website - About Zotero (2016) <https://www.zotero.org/about/>, [accessed on Feb 12].
- [7] Yamakawa, E. K., Kubota, F. I., Beuren, F. H., Scalvenzi, L. e Miguel, P. A. C. (may 2014). Comparativo dos softwares de gerenciamento de referências bibliográficas: Mendeley, EndNote e Zotero. *Transinformação*, v. 26, n. 2, p. 167–176.
- [8] Butros, A. e Taylor, S. (2010). Managing information: evaluating and selecting citation management software, a look at EndNote, RefWorks, Mendeley and Zotero. In *Netting knowledge: two hemispheres/one world: proceedings of the 36th IAMSLIC Annual Conference*.
- [9] Klock, A. C. T. e Gasparini, I. (2015). Sistemas de Gerenciamento de Referências: um comparativo dos aspectos colaborativos e sociais. In *Proceedings of the VI Workshop sobre Aspectos da Interação Humano-Computador na Web Social (WAIHCWS'15)*. SBC.
- [10] Moran, T. P. (1981). The Command Language Grammar: a representation for the user interface of interactive computer systems. *International Journal of Man-Machine Studies*, v. 15, n. 1, p. 3 – 50.
- [11] Barbosa, S. D. J. e Silva, B. S. Da (2011). *Interação Humano-Computador*. Rio de Janeiro: Elsevier Brasil.
- [12] Rogers, Y., Sharp, H. e Preece, J. (2011). *Interaction Design: Beyond Human - Computer Interaction*. Chichester: John Wiley & Sons Ltd.
- [13] Nielsen, J. (1993). *Usability Engineering*. San Diego: Academic Press Ltd.
- [14] Shneiderman, B. e Plaisant, C. (2005). *Designing the User Interface: Strategies for Effective Human-Computer Interaction*. New York: Addison-Wesley.
- [15] De Castro Salgado, L. C., Bim, S. A. e De Souza, C. S. (2006). Comparação entre os métodos de Avaliação de base Cognitiva e Semiótica. In *Proceedings of VII Brazilian Symposium on Human Factors in Computing Systems*. IHC '06. ACM.
- [16] Bastien, J. M. C. e Scapin, D. L. (1995). Evaluating a User Interface with Ergonomic Criteria. *Int. J. Hum.-Comput. Interact.*, v. 7, n. 2, p. 105–121.
- [17] Nielsen, J. e Molich, R. (1990). Heuristic Evaluation of User Interfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. , CHI '90. ACM. <http://doi.acm.org/10.1145/97243.97281>.
- [18] Prates, R.O. e Barbosa, S.D.J. (2003). Avaliação de Interfaces de Usuário–Conceitos e Métodos. *Jornada de Atualização em Informática do Congresso da Sociedade Brasileira de Computação*, Capítulo.