

Proposta de Metodologia para Estudo de Usuários da Embrapa



ISSN 1517-5111
ISSN online 2176-5081
Julho, 2012

*Empresa Brasileira de Pesquisa Agropecuária
Embrapa Cerrados
Ministério da Agricultura, Pecuária e Abastecimento*

Documentos 311

Proposta de Metodologia para Estudo de Usuários da Embrapa

*Maria Elisabeth Salviati
Francisco Eduardo de Castro Rocha*

Embrapa Cerrados
Planaltina, DF
2012

Exemplar desta publicação disponível gratuitamente no link:
http://bbeletronica.cpac.embrapa.br/versaomodelo/html/2012/doc/doc_311.shtml

Comitê de Publicações da Unidade

Presidente: *Cláudio Takao Karia*

Secretária-Executiva: *Marina de Fátima Vilela*

Secretárias: *Maria Edilva Nogueira*

Alessandra S. Gelape Faleiro

Supervisão editorial: *Jussara Flores de Oliveira Arbués*

Equipe de revisão: *Francisca Elijani do Nascimento*

Jussara Flores de Oliveira Arbués

Normalização bibliográfica: *Shirley da Luz Soares Araújo*

Editoração eletrônica: *Leila Sandra Gomes Alencar*

Capa: *Leila Sandra Gomes Alencar*

Ilustração da capa: *Wellington Cavalcanti*

Impressão e acabamento: *Divino Batista de Sousa*

Alexandre Moreira Veloso

1ª edição

1ª impressão (2012): tiragem 100 exemplares

Edição online (2012)

Todos os direitos reservados

A reprodução não autorizada desta publicação, no todo ou em parte, constitui violação dos direitos autorais (Lei no 9.610).

Dados Internacionais de Catalogação na Publicação (CIP) Embrapa Cerrados

S184p Salviati, Maria Elisabeth
Proposta de metodologia para estudo de usuários da Embrapa
/ Maria Elisabeth Salviati, Francisco Eduardo de Castro Rocha –
Planaltina, DF : Embrapa Cerrados, 2012.
67 p. — (Documentos / Embrapa Cerrados, ISSN 1517-5111,
ISSN online 2176-5081 ; 311).

1. Estudo de usuários-biblioteca. 2. Método quantitativo-
estatística. 3. Análise de conteúdo. 4. Metodologia. I. Rocha,
Francisco Eduardo de Castro. II. Título. III. Série.

020 - CDD 21

©Embrapa 2012

Autores

Maria Elisabeth Salviati

Bibliotecária, Ph.D. em Ciência da Informação,
Analista da Embrapa Cerrados,
elisabeth.salviati@embrapa.br

Francisco Eduardo de Castro Rocha

Engenheiro-agrícola, Psicólogo,
D.Sc. em Psicologia Social,
Pesquisador da Embrapa Cerrados,
francisco.rocha@embrapa.br

Sumário

Introdução.....	11
Usuários Internos da Biblioteca Eletrônica.....	15
Instrumento de coleta de dados	15
Escolha dos sujeitos	16
Entrevista	16
Análise dos resultados.....	17
Usuários Externos	46
Instrumento de coleta de dados	46
Escolha dos sujeitos	46
Análise dos resultados.....	47
Conclusão	52
Referências	55
Abstract	58
Anexo I. Roteiro da entrevista semiestruturada.....	59
Anexo II. Descrição dos resultados da Categoria III Acervo eletrônico.....	
Anexo III. Questionário aplicado aos usuários externos.....	66

Apresentação

Neste trabalho, descreve-se a metodologia utilizada no projeto de pesquisa do Macroprograma IV do Sistema Embrapa de Gestão, executado no período de setembro de 2007 a fevereiro de 2012 e intitulado “Estudo de Usuários das Bibliotecas Eletrônicas da Embrapa Cerrados, Embrapa Hortaliças, Embrapa Informação Tecnológica e Embrapa Recursos Genéticos e Biotecnologia”.

Ele aborda não apenas os aspectos teóricos da metodologia. Mas também descreve a metodologia em conjunto com a aplicação prática na qual foi utilizada, reportando-se com exemplos reais todos os aspectos metodológicos adotados, tais como: amostra, instrumentos de coleta de dados, operacionalização dos dados e sua análise.

A metodologia de análise de conteúdo conjugada utilizada em um dos estudos realizados já tinha sido aplicada, na Embrapa, para averiguar a opinião dos produtores de soja e dos extensionistas rurais, bem como a opinião dos produtores sobre o uso dos recursos hídricos. No entanto, acredita-se que ela possa ser aplicada em outras áreas dentro da Embrapa, em diversos processos que ofereçam serviços e produtos ao público usuário e a clientes, tais como: prospecção de demandas de clientes potenciais para um determinado produto a ser desenvolvido pelo processo de inovação tecnológica; avaliação do impacto de

tecnologias lançadas no mercado; e avaliação do próprio processo de transferência em relação ao seu objetivo principal, que é a adoção de tecnologias, produtos e serviços gerados ou adaptados pela Embrapa para benefício do cliente em seu sistema produtivo.

Por se tratar de uma metodologia que não tem tradição de uso na Embrapa, espera-se que esta proposta venha contribuir para outros estudos dessa natureza.

José Roberto Rodrigues Peres
Chefe-Geral da Embrapa Cerrados

Introdução

Estudo de usuários é uma linha de pesquisa clássica em Ciência da Informação que vem sendo empregada há mais de 40 anos com o objetivo de avaliar produtos e serviços de informação. Ao longo dos anos, foram utilizadas várias técnicas para coleta de dados. Os primeiros trabalhos utilizavam métodos estatísticos para medir a frequência de uso da informação. Foram estudos precursores da Bibliometria.

Nos últimos anos, entretanto, os estudos de usuários passaram a realizar uma investigação mais ampla – identificando as necessidades de informação dos usuários, que permitiu a adequação de sistemas e serviços às suas reais necessidades. Nesses estudos, é observado todo o comportamento do usuário na busca de informação, com o intuito de avaliar o seu nível de satisfação e o desempenho dos serviços de informação oferecidos.

Segundo definição de Figueiredo (1994), estudos de usuários são investigações que se fazem para saber o que os indivíduos precisam em matéria de informação, ou então, se suas necessidades de informação ao utilizarem uma biblioteca ou um centro de informação estão sendo satisfeitas de maneira adequada.

Sanz Casado (1994) define estudo de usuários como o *“conjunto de estudos que trata de analisar, qualitativa e quantitativamente, os hábitos de informação dos usuários, através da aplicação de diferentes métodos, entre estes os matemáticos, principalmente estatísticos, ao uso da informação”*.

Complementando, Costa et al. (2009) esclarecem que os estudos de usuário permitem verificar *Por que?, Como? e Para quais fins?* eles usam a informação e quais os fatores que afetam esse uso. Essas constatações contribuem para se conhecer as reais necessidades de informação dos usuários por meio da investigação de fatos da sua vida cotidiana. Os resultados desses estudos permitem oferecer serviços direcionados ao uso mais efetivo da informação.

Segundo Valls e Vergueiro (2006), ninguém é melhor que o próprio usuário para definir suas necessidades de informação e o nível de qualidade do serviço recebido.

Dessa maneira, esses estudos coletam dados diretamente com os usuários para criar e (ou) avaliar produtos e serviços de informação, bem como entender melhor o fluxo da transferência da informação (BAPTISTA; CUNHA, 2007).

A metodologia utiliza tanto métodos quantitativos quanto qualitativos. A pesquisa quantitativa caracteriza-se, tanto na fase de coleta de dados quanto no seu tratamento, pela utilização de métodos estatísticos. Já os métodos qualitativos permitem avaliar a interação do usuário em todo o processo de busca de informação. A combinação dessas abordagens em um único estudo ou em uma série de estudos (como no caso presente) permite identificar as necessidades individuais do usuário e implementar sistemas de informação adequados a essas necessidades (SANZ CASADO, 1994).

Para se estudar o comportamento humano, utilizam-se métodos qualitativos que procuram esclarecer quem são os usuários da informação, quais são suas necessidades e como procuram e usam a informação. Esses métodos levam em consideração também os aspectos da interação entre usuários e a tecnologia envolvida no serviço de informação.

Por sua vez, os métodos quantitativos caracterizam-se por possuir procedimentos claros e objetivos que enfatizam o desenvolvimento da investigação dentro de protocolos estabelecidos e técnicas específicas buscando expressar em termos de valores as necessidades dos usuários. Por isso, alguns autores consideram os resultados obtidos fortes em termos de validade externa, uma vez que podem ser generalizáveis para o conjunto da comunidade (COZBY, 2003).

No entanto, conforme afirmado por Baptista e Cunha (2007), alguns trabalhos de natureza totalmente quantitativa não conseguem almejar os objetivos desejados, principalmente devido à complexidade de se

determinar o comportamento e as necessidades de informação dos usuários, havendo necessidade de se utilizar, sempre que possível, os dois métodos integrados.

Nas pesquisas quantitativas, utilizam-se os questionários com perguntas fechadas e analisam-se os resultados contabilizando-se, estatisticamente, os resultados. Nas pesquisas qualitativas, os dados são coletados por meio de entrevistas ou questões abertas e pela observação. São, então, analisados utilizando-se técnicas interpretativas que visam descrever, traduzir e expressar o sentido dos fenômenos do mundo social (MINAYO, 1994).

Para se utilizar os métodos qualitativos, há necessidade de se buscar metodologias satisfatórias que trabalham em ambiente natural não controlado e que disponham de técnicas interpretativas adequadas para a avaliação de serviços de uma unidade de informação.

Novos métodos que cumprem esse papel surgiram dentro da pesquisa social. É o caso da análise de conteúdo, usada desde o início do século 20, com objetivos de medir a legibilidade de um texto ou comunicação e analisar atitudes, interesses e valores culturais de um grupo (BAPTISTA; CUNHA, 2007).

Para Bardin (2004), citado por Rocha et al. (2011b), a análise de conteúdo é um conjunto de procedimentos sistemáticos e objetivos de descrição do conteúdo das mensagens, visando obter indicadores que permitam a inferência de conhecimentos sobre o objeto de pesquisa.

A operacionalização desse tipo de análise é feita dividindo-se o texto de análise (ou corpus) em Unidades de Contexto Elementar (UCE). A UCE é definida como o menor segmento de um texto ou enunciado que têm uma significação em si (ROCHA et al., 2011a).

As fases subseqüentes incluem recorte manual das UCEs; aglomeração das UCEs que tratam de uma mesma temática e categorização das UCEs; operacionalização de categorias; e a inferência (RICHARDSON et al., 1999).

Esse método, no entanto, exige a realização de análise manual de palavras ou frases que ocorrem em um texto, seja ele de um documento ou da própria fala dos sujeitos. Por isso, pode até mesmo ser inviável em casos de grande volume de dados textuais. Para solucionar esse problema, surgiram os softwares que realizam parte dessas análises e simplificam o processo. O programa, porém, não realiza o detalhamento da análise. Por isso, seus resultados têm que ser conjugados com a análise manual que, nesse caso, é simplificada por ser realizada em parte dos dados (NASCIMENTO; MENANDRO, 2006).

Apresentam-se, a seguir, os procedimentos metodológicos utilizados para o desenvolvimento do projeto de pesquisa do Macroprograma IV do Sistema Embrapa de Gestão (SEG), desenvolvido no período de setembro de 2007 a fevereiro de 2012. Esse projeto teve por objetivo geral levantar, caracterizar e divulgar o estudo de usuários das Bibliotecas eletrônicas das Embrapa Cerrados, Embrapa Hortaliças, Embrapa Informação Tecnológica e Embrapa Recursos Genéticos e Biotecnologia.

Foram trabalhados, neste estudo, dois tipos de usuários das Bibliotecas eletrônicas supracitadas: interno e externo. Foram considerados usuários internos o público-alvo da biblioteca eletrônica constatado em estudo anterior (SALVIATI; CASTRO, 2009), formado por profissionais e estudantes de nível superior: pesquisadores, analistas, estagiários e bolsistas da Unidade da Embrapa investigada¹. Já o usuário externo é todo aquele de qualquer instituição (pública ou privada) no País ou no exterior e de qualquer profissão e nível educacional que acessaram a página remotamente.

Para análise dos usuários externos, foram utilizados quase que exclusivamente os métodos quantitativos devido às dificuldades de se realizar entrevistas com o público da Internet. A parte qualitativa da análise foi relativa a algumas questões abertas do questionário aplicado. Esses procedimentos encontram-se detalhados no tópico “Usuários externos da Biblioteca Eletrônica”.

¹ Existem Assistentes que realizam atividades de pesquisa. Por isso, cabe ao investigador determinar quais categorias devem ser inseridas no estudo, levando em consideração também o tempo para coleta e análise de dados que na pesquisa qualitativa são procedimentos complexos e morosos.

Por sua vez, utilizou-se a análise de conteúdo combinada para realizar a análise do usuário interno.

Usuários Internos da Biblioteca Eletrônica

A análise de conteúdo conjugada utiliza duas técnicas interpretativas do discurso originado pelas entrevistas: a análise de conteúdo de base semântica, segundo modelo Bardin (ROCHA et al., 2011a); e análise por meio do software Alceste de base lexical (Análise Lexical Contextual de um Conjunto de Segmentos de Texto) (SOUSA et al., 2009).

Essa metodologia foi empregada com o objetivo de se realizar uma análise detalhada do comportamento do usuário interno na consulta da Biblioteca Eletrônica, visando não só avaliar como o serviço está sendo utilizado, mas também obter direcionamentos para seu aprimoramento e desenvolvimento futuro.

Apresentam-se, a seguir, informações sobre o instrumento de coleta de dados, seleção da amostra, agendamento e aplicação da entrevista e procedimentos para a análise dos dados.

Instrumento de coleta de dados

Os usuários internos foram analisados por meio de uma entrevista semiestruturada, subdividida em duas fases. A Fase I continha 14 perguntas abertas e fechadas sobre o perfil do respondente: cargo, sexo, idade, escolaridade, experiência profissional, frequência de busca de informação pela internet, sites mais úteis para busca de informação, conhecimento e uso da Biblioteca Eletrônica, necessidades de divulgação e treinamento. A Fase II continha quatro perguntas, sendo três abertas e uma fechada sobre a Biblioteca Eletrônica com relação a: frequência de uso; opinião sobre os recursos de consulta utilizados, quanto à facilidade, rapidez e eficiência da busca; acervo, quanto à satisfação com o conteúdo e a necessidade de acréscimo de outros tipos de publicação; e recomendações. Apresenta-se, no Anexo I, o roteiro da entrevista semiestruturada.

Escolha dos sujeitos

O tamanho da amostra foi estabelecido de maneira não probabilística até a saturação das crenças, conforme sugerem os investigadores de pesquisas qualitativas (FONTANELLA et al., 2008; ROCHA et al., 2010; ROCHA et al., 2011a; ROCHA et al., 2011b). A priori, 16 sujeitos foram escolhidos, randomicamente, na lista de nomes de pesquisadores/analistas e estagiários/bolsistas fornecida pelo Setor de Gestão de Pessoas (SGP) de cada Unidade. Após a coleta de dados, examinaram-se as respostas para ver se havia pontos ainda a serem esclarecidos. Como se percebeu constante repetição das respostas em relação às questões apresentadas, encerrou-se a coleta de dados (ROCHA et al., 2010; ROCHA et al., 2011a; ROCHA et al., 2011b).

Embora o tamanho da amostra tenha sido determinado de maneira não probabilística, a escolha dos sujeitos foi realizada por amostragem probabilística estratificada e sistemática. Para escolha dos sujeitos de cada estrato (pesquisadores/analistas e bolsistas/estagiários), tomou-se o total de (empregado e não empregado) de cada categoria existente e dividiu-se por 8, eliminando-se os décimos. O valor encontrado foi o número de nomes que se contou nas listas do SGP para selecionar os 8 sujeitos entre pesquisador/analista e os 8 sujeitos entre estagiário/bolsista que seriam entrevistados, obedecendo-se rigorosamente a ordem.

Por exemplo: quadro de 95 empregados. Então: $95/8 = 11,875$ (valor aproximado para baixo é igual a 11). Elegeu-se a cada 11 pessoas da lista, um sujeito: 11°, 22°, 33°, 44°, 55°, 66°, 77° e 88°. Caso o empregado estivesse afastado por qualquer motivo, tomou-se o sujeito imediatamente seguinte (BARBETTA, 2002).

Entrevista

Após o sorteio dos indivíduos, houve comunicação com os sorteados por telefone, esclarecendo o objetivo do trabalho. Nessa oportunidade, foram efetuadas 14 perguntas abertas e fechadas da Fase I sobre o perfil do respondente. As respostas foram registradas à caneta no questionário impresso. Agendaram-se as entrevistas da Fase II apenas para os

sujeitos que já conheciam e tinham consultado a Biblioteca Eletrônica, isto é, quando o sujeito respondia (questão 10) que não conhecia a Biblioteca Eletrônica ou que nunca tinha consultado a Biblioteca Eletrônica (questão 11), escolhia-se o nome imediatamente seguinte.

No dia da realização da entrevista, houve nova comunicação com os usuários selecionados. Esse procedimento evitou em grande parte o não comparecimento e, conseqüentemente, o atraso da coleta de dados.

O local escolhido para a sua realização foi a Biblioteca da Unidade. Lá, ficou disponível um computador com acesso à Biblioteca Eletrônica para que o entrevistado pudesse examiná-la com mais detalhe para responder questões específicas sobre ela, quando sentisse necessidade.

A entrevista (Fase II) foi registrada com gravador de voz digital. Gravaram-se as perguntas, as respostas e até mesmo a sessão de consulta à Biblioteca Eletrônica, quando realizada.

O entrevistador era Bibliotecário, conhecedor da Biblioteca Eletrônica e das questões. Ele estava munido do questionário e lia cada questão para o sujeito. Por tratar-se de uma entrevista e não de um questionário, as perguntas não foram entregues para o sujeito respondê-las sozinho.

Terminada a entrevista, a gravação era transferida para o computador e, então, transcrita na íntegra (perguntas e respostas) e sempre que possível, no mesmo dia, utilizando-se o editor de texto Word XP.

Análise dos resultados

As perguntas fechadas sobre o perfil do entrevistado, oriundas da Fase I da entrevista, foram analisadas em termos de frequência e porcentagens, utilizando-se o aplicativo Excel. Elas foram úteis para mostrar as características biodemográficas dos respondentes, tais como, cargo ocupado, sexo, idade, nível de escolaridade, experiência profissional, frequência de busca de informação pela internet, sites

mais úteis para busca de informação, conhecimento e uso da Biblioteca Eletrônica, necessidades de divulgação e treinamento. Por exemplo, na Tabela 1, mostra-se o resultado para Escolaridade.

Tabela 1. Dados biodemográficos dos entrevistados: Escolaridade.

Categoria	Subcategoria	F	%
Escolaridade	Doutorado	9	56,25
	Mestrado	4	25
	Especialização	2	12,50
	Ensino superior	1	6,25
	Pós-doutorado	-	-

Por sua vez, as perguntas da Fase II, relativas especificamente ao uso da Biblioteca Eletrônica, foram analisadas com a metodologia de Análise de Conteúdo combinada.

Os procedimentos utilizados para a análise dos dados podem ser subdivididos em três fases:

1. Análise lexical: realizada por meio do software Alceste.
2. Análise de conteúdo de Bardin.
3. Análise das relações existentes entre as classes.

Análise lexical

Essa fase envolveu várias etapas que serão descritas em detalhes nos itens a seguir: preparação do banco de dados *corpus* e sua submissão ao *software* Alceste.

O Alceste (Analyse Lexicale par Context d'un Ensemble de Segments de Texte)² é um software para análise de dados textual do Centre National de la Recherche Scientifique (CNRS), criado por Max Reinert, em 1986, na França e de propriedade da Image Informatique Mathématiques Gestion. É uma ferramenta fundamental para auxiliar a análise de dados textual e aplicável a dados provenientes de

² Análise lexical vem da palavra léxico, do grego Lexicom, é sinônimo de vocabulário. É o conjunto de vocábulos de forte conteúdo semântico, tais como substantivos, adjetivos, verbos e advérbios de modo que co-ocorrem/aparecem simultaneamente juntos em trechos no texto (CARVALHO, 2003)

questões abertas, obras literárias, artigos de revistas, ensaios, livros e documentos diversos.

Ele tem por objetivo descrever, classificar, assimilar e sintetizar, automaticamente, um texto, utilizando o método de Classificação Descendente Hierárquica. Ele quantifica um texto extraindo as estruturas mais significativas. Essas estruturas estão intimamente relacionadas com a distribuição de palavras do texto, que raramente é aleatória.

O método realiza sucessivos fracionamentos do texto em vocábulos significativos, a partir dos valores de *chi-quadrado* que eles possuem. Em seguida, identifica as oposições mais fortes entre as palavras do texto e extrai as classes de declarações representativas. Finalmente, o sistema representa graficamente as classes principais do discurso identificadas pela Análise Fatorial por Correspondências.

O software Alceste encontra suas aplicações nas áreas de Sociologia, Psicologia, Análise do Discurso, Marketing, Publicidade, Jornalismo, História, Linguística, além de outras que possuem grandes quantidades de texto a ser processado. Ele lida com todos os tipos de texto eletrônicos. É executado em ambientes Windows XP, Vista, 7 e Mac OS X (em computadores com um processador Intel e Windows).

Preparação do corpus

Para realizar essa análise, todas as entrevistas já transcritas em Word foram copiadas em um novo arquivo e reformatadas conforme padrão solicitado pelo aplicativo. Guardaram-se, porém, os arquivos originais para serem utilizados durante a análise de conteúdo de Bardin.

A preparação do *corpus* para a análise do Alceste envolveu a eliminação de todas as perguntas e falas relativas ao entrevistador. O texto resultante, então, foi corrigido obedecendo-se as seguintes regras adaptadas de Sousa et al. (2009):

(a) Eliminação dos seguintes sinais:

- Aspas
- Apóstrofo

- Cifrão
 - Porcentagem
 - Asterisco
 - Interrogação
 - Exclamação
 - Reticências
 - Travessão
 - Negrito, itálico, grifo e outros sinais similares
 - Recuo de parágrafo, margens ou tabulações do texto
 - Justificação do texto
- b) Utilização de apenas os seguintes sinais de pontuação:
- ponto
 - ponto e vírgula
 - vírgula
- c) Formatação de texto todo corrido, sem mudança de linha.
- d) Uso de maiúsculas só para nomes próprios.
- e) União das palavras compostas com *underline*. Ex.: Distrito_Federal.
- f) Troca do hífen das palavras compostas por *underline*.
- g) Padronização das siglas e nomes próprios para obedecer sempre a mesma grafia.
- h) Revisão gramatical do português, corrigindo grafia e concordância.
- i) Complementação de todas as frases incompletas: cada frase deve encerrar um sentido completo e não deve possuir palavras subtendidas. Complementar com as palavras necessárias, sem modificar o sentido. Se necessário, reexaminar o texto original para escolher as palavras adequadas. Caso haja impossibilidade de completar determinadas frases, elas deverão ser eliminadas.

- j) Eliminação de expressões sem necessidade, tais como: Ahh, Uhmm, etc.
- k) Eliminação de frases não condizentes com o assunto tratado.

Na Figura 1, mostra-se o trecho original de uma entrevista e, na Figura 2, traz-se o mesmo trecho no *corpus* preparado.

Com que frequência você utiliza, olha temos as opções: diária, semanal, mensal, esporádica.

Isso varia muito. Se eu to fazendo um projeto, às vezes, diária... mas às vezes também...

Na média, você poderia dizer o que?

Eu poderia colocar semanal, mas, obviamente, não é toda semana que eu consulto.

Entendi. É na média né?!

Mas outras vezes é diária.

Eu tenho aqui diversos tipos de consulta: por assunto, por autor, tipo, título e consulta avançada. Você já utilizou todas essas formas, ou não?

Não. Eu uso a por assunto e por autor.

É?

Normalmente.

Figura 1. Trecho original transcrito de uma entrevista realizada.

A frequência de uso da biblioteca eletrônica varia muito. Se eu estou fazendo um projeto, às vezes, a frequência de uso é diária. Eu poderia colocar frequência de uso semanal, mas, obviamente, não é toda semana que eu consulto. Mas outras vezes a frequência de uso é diária. Não utilizo todas as formas de consulta. Eu uso a consulta por assunto e por autor, normalmente.

Figura 2. Mesmo trecho anterior preparado para o *corpus* de análise automática.

Em seguida, todas as entrevistas foram reunidas em um só arquivo Word, contendo uma linha de comando para separá-las, isto é, a parte do *corpus* relativa a cada sujeito inicia um novo parágrafo que é precedido pela linha de comando. A linha de comando identifica cada sujeito, conforme as variáveis biodemográficas que especificamos para classificá-lo. Essas variáveis têm que caber em apenas uma linha de texto, por isso devem ser abreviadas (SOUSA et al., 2009); elas têm uma sintaxe própria para serem entendidas pelo programa: quatro asteriscos identificam um novo sujeito, por isso é comumente chamada linha estrela. Cada variável é identificada por um asterisco. Um sinal de *underline* separa o nome da variável de seu valor.

Apresentam-se, na Tabela 2, as variáveis utilizadas especificamente neste estudo e, na Figura 3, a linha de comando construída para o sujeito 1.

Tabela 2. Exemplo de variáveis da linha de comando.

Variável	Variável padronizada	Categoria
Sujeito	subj	Numeração de 01 a 16
Unidade	uni	Unidade: cpac , cenargen , cnph , sede
Cargo	car	Cargo: pesq (<i>pesquisador</i>), est (<i>estagiário</i>), bol (<i>bolsista</i>), ana (<i>analista</i>)
Sexo	sex	Sexo: mas (<i>masculino</i>), fem (<i>feminino</i>)
Idade	ida	Idade: m30 (menos de 30), 30-39 , 40-49 , 50-59 , 60m (60 ou mais)
Escolaridade	esc	Escolaridade: phd (<i>doutorado</i>), ms (<i>mestrado</i>), sup (<i>superior</i>), esp (<i>especialização</i>), pdr (<i>pós-doctor</i>)
Anos de experiência na Embrapa	anos	Anos de Embrapa: m2 (<i>menos de 2</i>), 2-5 , 6-10 , 11-15 , 15m (<i>mais de 15</i>)
Anos de experiência fora da Embrapa	fora	Anos fora da Embrapa: m2 (<i>menos de 2</i>), 2-5 , 6-10 , 11-15 , 15m (<i>mais de 15</i>)
Frequência de uso da Internet	int	Frequência de uso da internet: dia (<i>diária</i>), sem (<i>semanal</i>), mes (<i>mensal</i>), esp (<i>esporádica</i>)

```
*** *suj_01 *uni_cpac *car_pesq *sex_mas *ida_40-49 *esc_phd  
*anos_2-5 *fora_15m *int_mes
```

Figura 3. Linha de comando do sujeito 1.

As variáveis com valores em branco (quando o usuário não respondeu a pergunta), receberam valor *sem*.

O arquivo assim preparado no Word deve ser gravado com formato “texto sem formatação” (extensão txt), incluindo o retorno de carro (alimentação de linha). O nome desse arquivo não pode ultrapassar sete caracteres e nem possuir espaços em branco ou sinal de *underline* (SOUSA et al., 2009).

Submissão do corpus ao software Alceste

Os passos para a realização da análise lexical pelo software Alceste envolveu: configuração do aplicativo, interpretação dos resultados e elaboração do dendograma.

Apresentam-se, a seguir, algumas instruções de como utilizar o aplicativo adequadamente. Mais informação poderá ser obtida no trabalho de Sousa et al. (2009).

Primeiramente, deve-se iniciar o Alceste e escolher *Traitement/Nouveau*, e, em seguida, escolher a pasta e o arquivo com extensão txt que contém o corpus para análise.

O sistema apresenta, em uma nova janela, o *corpus* aberto, indicando a língua do dicionário que será utilizado. Deve-se, nesse momento, escolher a língua portuguesa, caso esta não esteja indicada.

Logo abaixo do dicionário, existem as seguintes opções: *Analyser le corpus* e *Paramétrage*. Na primeira opção, o sistema faz uma varredura no corpus para buscar possíveis erros de formatação. Deve-se usar sempre essa opção quando não se tem certeza se o corpus está correto. Se houver erros, aconselha-se examinar atentamente o *corpus*

e providenciar a sua correção antes de prosseguir. Caso não haja erros, o sistema inicia a análise do *corpus*, tomando como parâmetros aqueles definidos por *default* (padrão). É na opção *Paramétrage* que se faz a configuração do aplicativo e, portanto, deve ser a primeira a ser escolhida. O guia do Alceste preparado por Sousa et al. (2009) mostra, passo a passo, cada tela da utilização do Alceste em uma análise de dados textual.

Configuração do aplicativo

Na opção *Paramétrage*, é possível configurar o aplicativo de acordo com os parâmetros que devem ser observados durante a análise. Ao escolher essa segunda opção, o sistema apresenta a tela de *Paramétrage simplifié* que possui quatro etapas numeradas de A até D e que apresentam algumas opções, conforme a seguir. Essa configuração é simples. Existe, também, a configuração avançada que só é aconselhável para usuários experientes.

Etapa A: *Lecture du texte et calcul des dictionnaires*

Nessa etapa, devem ser selecionadas as opções *Lemmatisation* e *Prendre les paragraphes du texte comme UCE*. O programa contabiliza a frequência das palavras dentro do corpus (discurso). A opção de *Lemmatisation* permite considerar a frequência das palavras a partir de sua raiz, de forma que as palavras com mesma raiz, tais como trabalh (trabalha, trabalham, trabalhando, trabalhar, trabalho, trabalhoso) e consult (consulta, consultado, consultam, consultando, consultar), sejam contabilizadas como palavras únicas, isto é, com mesmo sentido. Isso é muito importante para a análise porque aumenta a frequência de conteúdos significativos no universo dos sujeitos.

A segunda opção, *Prendre les paragraphes du texte comme UCE*, permite definir cada frase como a unidade de contexto elementar. Ela é definida pelo Alceste, baseada nos sinais de pontuação e no número de palavras.

Etapa B: *Définition des UCE et classification*

Nessa etapa, deve-se marcar a opção *Classification simple sur les*

unités de contexte élémentaires (UCE). Dependendo do tamanho do corpus, pode-se requerer a classificação dupla de cada UCE. A classificação das unidades de contexto é realizada em função da distribuição do vocabulário e pode ser simples ou dupla, conforme o contexto.

A classificação simples utiliza a técnica de classificação descendente hierárquica desenvolvida para o tratamento de matrizes esparsas. É realizada sobre as unidades naturais, isto é, divisões naturais do discurso: capítulos, seções, itens ou respostas de cada entrevista realizada.

Já a classificação dupla executa duas classificações sucessivas e é utilizada quando existem unidades de contexto de diferentes comprimentos, visando manter a estabilidade das classes (LEBART; SALEM, 1994; ALCESTE, s.d.).

No nosso caso, a classificação simples foi suficiente.

Etapa C: *Définition des classes et calcul de l'A.F.C.*

Nessa etapa, deve-se marcar a opção *Calcul automatique du nombre minimum d'UCE* por classe, deixando o programa escolher quantas UCEs podem pertencer a cada classe. As classes são os temas principais do discurso identificados pelo programa a partir da coocorrência de palavras.

Etapa D: *Calculs complémentaires*

Nessa etapa, devem-se marcar todas as três opções:

- Liste des UCE significatives par classe: o sistema lista todas as UCE significativas encontradas em cada classe.
- Recherche des segments repetes: pesquisa segmentos de texto repetidos.
- Classification Ascendante Hiérarchique: realiza a classificação ascendente hierárquica das classes, mostrando como elas estão inter-relacionadas.

Interpretação dos resultados

Após a análise do corpus (16 entrevistas), o software Alceste apontou o total de categorias (classes) existentes no discurso; o total e porcentagem de UCEs existentes e analisadas no corpus; e o total de UCEs e sua porcentagem para cada categoria.

O aplicativo apresentou, ainda, uma lista das raízes lexicais de cada categoria, contendo sua frequência de ocorrência, χ^2 (qui-quadrado) e porcentagem de UCEs que continham aquela raiz dentro da categoria.

A análise estatística do qui-quadrado encontra um valor da dispersão para duas variáveis nominais, avaliando a associação existente entre variáveis qualitativas. Trata-se de um teste estatístico não paramétrico, ou seja, não depende dos parâmetros da população, com a média e a variância. É um teste que compara proporções, isto é, as possíveis divergências entre as frequências observadas e esperadas para um certo evento, no caso a coocorrência dos léxicos. O Alceste aplica esse teste automaticamente.

A partir da análise fornecida pelo Alceste, foi possível aprofundar e detalhar o estudo (passo 2) por meio de análise semântica das subcategorias existentes em cada categoria encontrada. Esse método, chamado de análise de conteúdo de Bardin, citado por Rocha et al. (2011b), foi realizado em 50% das entrevistas, escolhidas aleatoriamente, e que levantou as opiniões semelhantes e divergentes entre elas. Esse procedimento é descrito neste trabalho, no tópico “Análise de conteúdo, modelo Bardin.”

Elaboração do dendograma

A partir dos dados extraídos do relatório Rapport d’analyse: résultats généraux, preparou-se o Dendograma (representação diagramática ramificada da inter-relação entre as classes do discurso), conforme Figura 4. Para elaborá-lo, tomou-se a *Arbre de classification* descendente, a fim de definir a estrutura do dendograma – hierarquia e inter-relacionamento entre as classes. O programa pode apresentar até 10 classes (default) cuja hierarquia pode variar.

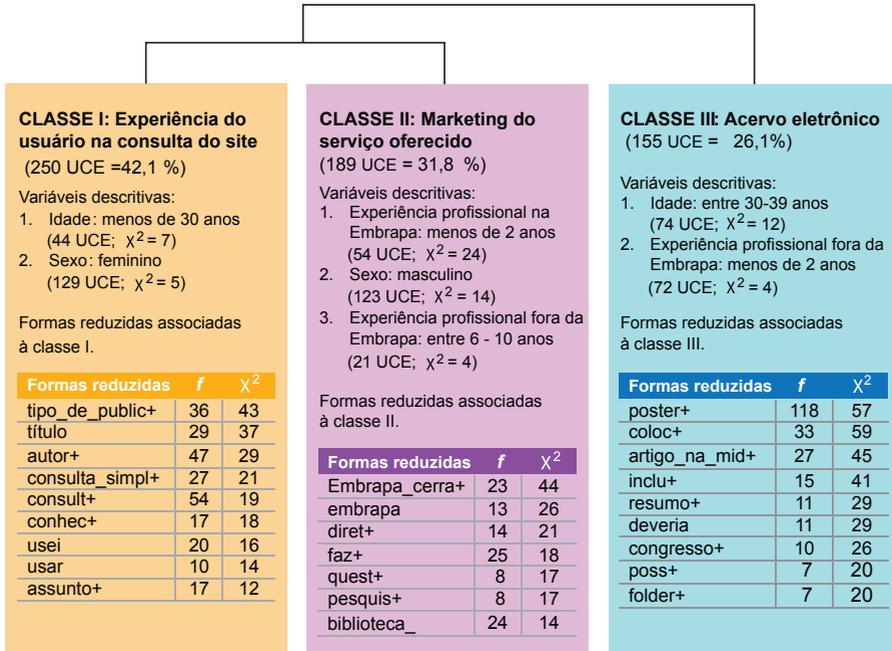


Figura 4. Distribuição de classes relativas ao uso da biblioteca eletrônica da Embrapa Cerrados pelos usuários internos (N = 16).

Apresentam-se, nas Figuras 5, 6 e 7, exemplos de estruturas de árvore possíveis e os comentários sobre o tipo de inter-relacionamento que indicam, a fim de orientar o usuário na interpretação adequada do gráfico.

Pode-se ver, na Figura 5, uma árvore que possui três classes. A maior classe é, portanto, a mais importante é a Classe 3, que possui 41% das UCEs significativas. A classe 2 aparece em segundo lugar com 33,3% das UCEs e a classe 1, com 25,6% das UCEs. Entre as classes 2 e 3, existe um relacionamento mais estreito, pois estão ligadas hierarquicamente em mesmo nível. Já comparando essas classes com a classe 1, o relacionamento existente é mais distante.

Já na Figura 6, nota-se a existência de quatro classes agrupadas duas a duas. A classe 1 e 4 estão mais próximas, assim como as classes 2 e 3 também possuem maior proximidade entre si. No entanto, os dois grupos

possuem relacionamentos mais distantes entre si. A maior classe é a 3, com 40% das UCEs. Em seguida, aparece a classe 2, com 25%. Esse grupo domina quase todo o discurso, com 65% das UCEs. As classes 1 e 4, pertencentes a outro grupo hierárquico com menores percentuais de UCEs: 14% e 21% respectivamente, perfazem o total de 35% das UCEs significativas. São, portanto, classes do discurso menos dominante.

Première classification descendante

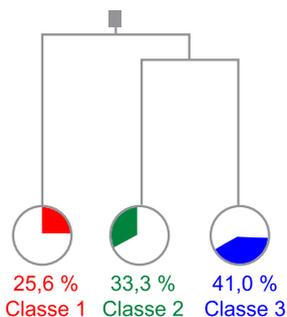


Figura 5. Árvore de classificação ascendente: exemplo 1.

Première classification descendante

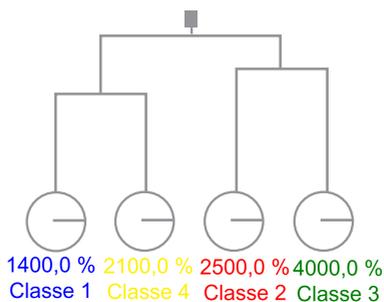


Figura 6. Árvore de classificação ascendente: exemplo 2.

Verifica-se, na Figura 7, uma árvore com três classes e todas de mesmo nível hierárquico, mostrando haver inter-relacionamento de mesmo nível entre elas. A maior classe é a 2, com 46% das UCEs, seguida pela classe 3 com 35%. A classe 1 possui o menor número de UCEs, apenas 19%.

Première classification descendante

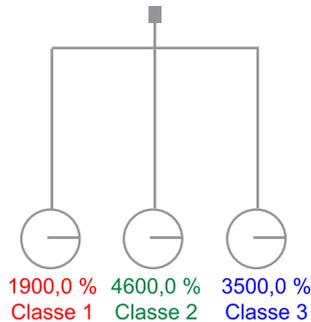


Figura 7 . Árvore de classificação ascendente: exemplo 3.

A estrutura hierárquica das classes criada pelo sistema Alceste deve ser representada no dendograma. Esse gráfico possui, ainda, três tipos de informações: (a) o título de cada uma das classes, atribuído pelo pesquisador com base no tema do discurso relativo àquela classe e seu total e percentual de UCE; (b) as variáveis descritivas relativas ao perfil que são significativas na classe; e (c) as unidades lexicais predominantes na classe, com sua frequência (f) e χ^2 .

A) Título de cada uma das classes

Deve-se nomear cada uma das classes com base no conteúdo do discurso que ela estiver representando, indicando o total de UCEs e sua porcentagem relativa ao todo do discurso. As frequências e porcentagens de UCEs de cada classe são obtidas no gráfico denominado *Répartition des UCE classées*, apresentado no relatório emitido pelo Alceste (Figura 8).

Já para se nomear as classes, é necessário, primeiramente, identificar o assunto principal do discurso de cada classe, e, então, escolher um título apropriado. Para isso, consultam-se as palavras significativas de cada classe, mostradas no relatório do Alceste no item denominado *Résultats de la classe* nº 1, 2, 3 e etc (Figura 9).

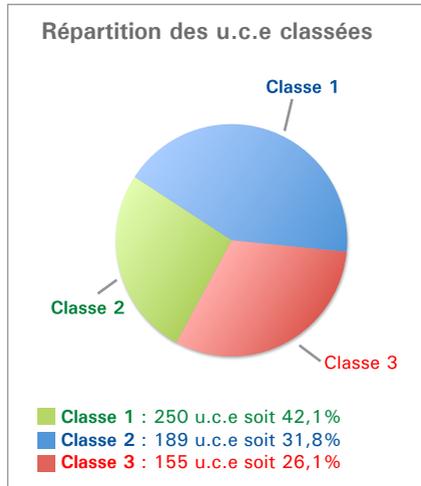


Figura 8. Classificação do discurso.

Résultats de la classe n°1

éssences significatives

Voici le vocabulaire caractéristique (formes réduites) de la classe 1 en fonction du kh2.
 Effectif1 : effectif réel du mot dans la classe ;
 Effectif2 : nombre d'u.c.e de la classe contenant le mot ;
 Total : nombre total d'u.c.e classées contenant le mot ;
 Percent : pourcentage du nombre d'u.c.e de la classe contenant le mot.

Vocabulaire	Kh2	Effectif 1	Effectif 2	Total	Percent	Catégorie Grammaticale
por	49	86	86	123	70%	Prépositions simples et
lipo_de_publico	43	36	32	33	97%	Formes non reconnues
titulo	37	29	26	26	100%	Formes non reconnues
vez	29	34	33	40	83%	Formes non reconnues
autor	29	47	39	50	78%	Formes non reconnues
quer	27	33	27	31	87%	Formes non reconnues
as	25	59	52	76	66%	Formes non reconnues
vou	22	26	23	27	85%	Formes non reconnues
consulta_simpl	21	27	26	32	81%	Formes non reconnues
consult	19	54	45	67	67%	Formes non reconnues
conhec	18	17	13	13	100%	Formes non reconnues
usei	16	20	17	20	85%	Formes non reconnues
usar	14	10	10	10	100%	Formes non reconnues
das	13	13	13	15	87%	Formes non reconnues
's uj_12	12	17	17	22	77%	Formes non reconnues
ass unto	12	17	16	20	80%	Formes non reconnues
rab	11	34	31	48	65%	Formes non reconnues
erman	11	8	8	3	100%	Formes non reconnues
dificuldade	11	13	12	14	86%	Formes non reconnues
pod	10	17	17	23	74%	Formes non reconnues
consulta_avan	10	17	17	23	74%	Formes non reconnues
arvore_hiperbo	10	18	14	18	78%	Formes non reconnues
estrutura_de_a	9	21	19	27	70%	Formes non reconnues
ou	8	21	21	32	66%	Conjonctions et locutions
tema	8	9	8	3	89%	Formes non reconnues

Figura 9. Palavras significativas da Classe 1.

Na Figura 9, apresenta-se cada palavra significativa encontrada na classe 1, por ordem de importância denotada pelo qui-quadrado (χ^2). As colunas seguintes se referem: Effectif1 (força da palavra na classe); Effectif2 (frequência de UCE que contém a palavra na classe); Total (total de UCE do corpus que contém a palavra); e porcentagem (relativa à frequência de UCE que contém a palavra na classe).

Deve-se consultar, também, a relação de UCEs significativas de cada classe, apresentada no relatório sob o título de Unités de Contexte Élémentaires de la classe 1, 2, 3 e etc., conforme Figura 10.

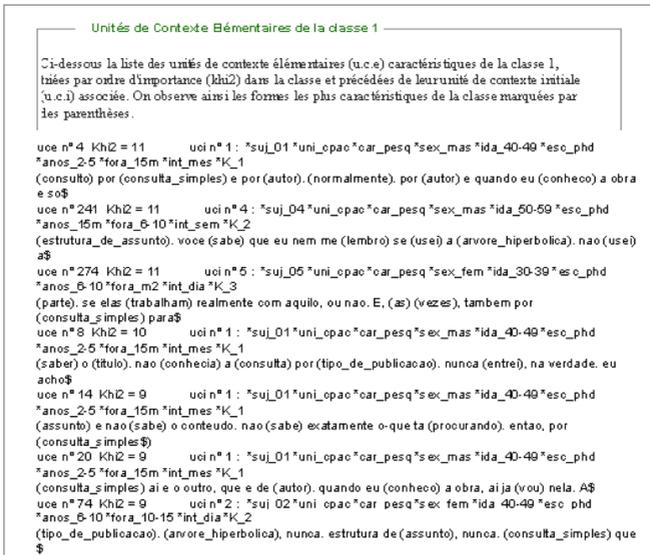


Figura 10. UCEs significativas da Classe 1.

Na Figura 10, mostram-se as UCEs com as palavras significativas no contexto, isto é, onde foram encontradas. Essas frases nos permitem identificar os assuntos que foram abordados na classe e a partir deles nomear adequadamente cada classe.

B) Indicação das variáveis descritivas da classe

As variáveis descritivas são aquelas que entraram na linha de comando antes do discurso de cada sujeito. O programa analisa essas variáveis

indicando quais foram os sujeitos predominantes em cada classe. Para identificá-las, consulta-se a lista de palavras significativas da Figura 9. Essa lista traz as variáveis da classe, facilmente identificáveis porque possuem o asterisco precedendo seu nome. Selecionam-se aquelas significativas, isto é, que possuem qui-quadrado maior ou igual a 4. Junto com essas variáveis, indicam-se a frequência de UCEs em que aparecem e seu qui-quadrado.

C) Lista das unidades lexicais predominantes na classe

A última parte do dendograma contém a lista das palavras significativas (raízes) encontradas, sua frequência dentro da classe e o respectivo qui-quadrado. O sinal + indica que a palavra foi truncada para pegar todas as terminações. Selecionam-se essas palavras a partir do relatório, seção de *Résultats de la classe* (Figura 9); escolhem-se as mais significativas para o tema do discurso (classe) em questão, desprezando-se palavras com qui-quadrado inferior a 4.

O dendograma é fundamental para se realizar a análise manual detalhada, etapa 2 a seguir.

Análise de conteúdo de Bardin

A análise de conteúdo, modelo Bardin, é uma análise manual, cujo objetivo é detalhar o resultado obtido na análise automática. Esse detalhamento é efetuado sobre parte do discurso, geralmente 50% da amostra.

Resumidamente, o procedimento consiste em se examinar, minuciosamente, as classes identificadas pelo Alceste, levantando-se todas as UCEs provenientes do discurso de parte dos entrevistados selecionados. Essas UCEs são classificadas em categorias e subcategorias de assunto e, posteriormente, organizadas hierarquicamente e encaixadas dentro das classes do Alceste. As quantidades (frequências) de UCEs e porcentagens são calculadas para cada subcategoria e categoria e geram a tabela hierárquica de frequências de UCEs. É por meio dela que se interpretam e se redigem os resultados.

Por se tratar de uma análise manual, ela é bastante trabalhosa e demanda muito tempo para sua conclusão. Porém, quando se usa a metodologia combinada, Alceste e Bardin, a análise de conteúdo de Bardin é simplificada e realizada somente sobre parte da amostra.

No trabalho citado de exemplo da metodologia, selecionaram-se 50% da amostra: 4 sujeitos com cargo de analista/pesquisador e 4 sujeitos com cargo de bolsista/estagiário. Apresentam-se, a seguir e em detalhe, todas as etapas seguidas para a análise pelo modelo Bardin:

- 1) Preparação do corpus para análise manual.
- 2) Classificação das UCE em categorias e subcategorias.
- 3) Separação das categorias e subcategorias conforme as classes do Alceste.
- 4) Elaboração da tabela de categorização de UCE contendo frequência e porcentagem de UCE por cada categoria e subcategoria.
- 5) Interpretação e redação dos resultados.

Preparação do corpus para análise manual

Para a realização da análise automática, o corpus precisa obedecer a certas regras para ser compreendido pelo programa. Já na análise manual, é necessário realizar uma adaptação do corpus para permitir a compreensão perfeita das UCEs. É a partir das UCEs que toda a análise será efetuada.

Utilizam-se os arquivos originais de transcrição das entrevistas (cada entrevista gera um arquivo em Word que contém as perguntas e respostas) para fazer essa adaptação. O processo de adaptação é efetuado mantendo-se os arquivos individuais até o processo de classificação de todas as UCEs, quando todo o discurso sob análise é reunido em um único arquivo Word. Os passos são os seguintes:

a) Complementação do sentido de todas as frases

Todas as frases provenientes do entrevistado devem ser completadas sempre que possível, caso contrário, devem ser eliminadas. É importante observar que esse acréscimo deve respeitar o discurso do participante, isto é, esclarecer o que foi dito sem alterar o seu sentido. Para esse processo, pode-se levar em conta também as falas do entrevistador. Toda palavra acrescentada deve vir entre parênteses no corpus.

b) Eliminação das perguntas

Eliminar todas as perguntas e a conversação do entrevistador.

c) Eliminação de palavras sem necessidade, tais como: Ahh, Uhhh

Ao eliminá-las, em seu lugar, deverá ser colocado o sinal de reticências entre parênteses. Ex.: (...).

d) Revisão gramatical do português corrigindo grafia e concordância

A revisão gramatical visa permitir que a análise seja efetuada corretamente. Alguns erros de pontuação, grafia, etc podem prejudicar a interpretação do discurso.

e) Particionamento do texto em (UCEs)

A separação de UCEs, isto é, o texto em frases de sentido completo, não é um procedimento simples. Para se separar os períodos compostos é necessário realizar análise sintática, que objetiva examinar a estrutura de um período e das orações que o compõem para identificar orações coordenadas e subordinadas. Enquanto as orações coordenadas devem ser separadas, formando várias UCEs, as subordinadas são mantidas juntas, formando uma única UCE. A divisão correta das UCEs é importante porque ela será contabilizada e o número de UCEs de cada categoria e subcategoria determinará a importância do discurso naquela categoria e subcategoria. Para detalhes desse procedimento, consulte Rocha et al. (2011a).

f) Eliminar frases não condizentes com o assunto tratado

Às vezes, o entrevistado não responde sobre o que lhe é perguntado e aparecem no texto frases que nada tem a ver com a entrevista. Nesse caso, essas frases devem ser eliminadas do *corpus*.

Apresentam-se, nas Figuras 11 e 12, trechos de uma entrevista original e adaptada.

Você acha que mais alguma publicação que esteja faltando, que seja da Embrapa Cerrados? São esses mesmo né?

Uma coisa que eu acho que não deveria colocar, não sei se ainda tem, mas Poster de Trabalho. Eu acho que agora não se coloca mais né?!

Aham

Eu não acho que é muito legal colocar poster, porque a informação é bem... muitas vezes, que nem... não vou dizer que é todo mundo, mas muita gente coloca tópicos, então eu acho que não é uma coisa com muita serventia. E, depois, você bota ali e, dependendo do que tem no poster, pode até te prejudicar na hora de publicar né?! A pessoa olha lá e está disponível, por exemplo, na internet a informação e você pega, vai lá e já publicou uma informação que não é nem uma revista, nem nada. É só um poster que está na internet, sabe? Eu acho que isso não deveria ser colocado né?! Eu acho que não está sendo mais né?

Figura 11. Trecho original de uma entrevista.

não sei se ainda tem (pôster de trabalho)/[11]

mas, uma coisa que eu acho que não deveria colocar (...) (é) pôster de trabalho/[11]

Eu acho que agora não se coloca mais (pôster) /[11]

Eu não acho que é muito bom colocar pôster/[11]

muitas vezes, (...) não vou dizer que é todo mundo, mas muita gente coloca tópicos (nos pôsteres)/[11]

eu acho que (pôster) não é uma coisa com muita serventia/[11]

dependendo do que tem no pôster, (se você divulga) pode até te prejudicar na hora de publicar/[11]

A pessoa olha lá e está disponível, por exemplo, na internet a informação (pôster)/[11]

você pega (...) uma informação que não é nem uma revista, nem nada/[11]

É só um pôster que está na internet/[11]

Eu acho que (...) (pôster) não deveria ser colocado/[11]

Eu acho que (pôster) não está sendo mais (colocado) /[11]

Figura 12. Mesmo trecho da entrevista adaptada para a análise de conteúdo.

No texto adaptado, notam-se frases curtas, limpas, de sentido completo e português correto. Essas frases tiveram palavras acrescentadas e outras eliminadas, mas preservaram o sentido original. Cada frase recebeu um código final que identificou o indivíduo que a proferiu. Esse procedimento facilita a busca do texto original quando é necessário.

Classificação das UCEs em categorias e subcategorias

Depois que todas as entrevistas foram revistas, conforme instruções no passo 1, realiza-se a classificação das UCEs em categorias e subcategorias. A categorização não é difícil de ser preparada. Categorizar UCEs significa classificá-las e agrupá-las por assunto, obedecendo à ordem hierárquica. O ponto de partida é a classificação inicial fornecida pelo Alceste. As classes primárias de assunto (aquelas de hierarquia mais alta) foram determinadas pelo software Alceste, e utilizadas na construção do dendograma.

Nessa etapa, as classes do Alceste devem ser utilizadas para separar em classes principais todas as UCEs semânticas. Para efetuar esse procedimento, lança-se mão do dendograma, do relatório do Alceste e dos arquivos em Word preparados na etapa anterior. Preparam-se, então, novos arquivos, nos quais são colocados os títulos das classes primárias identificadas pelo Alceste. Em baixo de cada classe, reúnem-se todas as UCEs que pertencem a ela.

Para identificar as UCEs de cada classe, primeiramente, identifica-se o assunto da UCE e, então, confrontando-se com os títulos das classes, raízes do dendograma e UCEs lexicais listadas no relatório do Alceste, define-se a qual classe ela pertence. O processo resumido pode ser descrito como identificar o assunto de cada UCE e movê-la para a classe correspondente.

Em seguida, as UCEs, já separadas em classes primárias, serão reclassificadas em subcategorias primárias, secundárias e terciárias. Para esse procedimento, agrupam-se as UCEs de mesma classe em subgrupos de assuntos mais específicos em hierarquias dependendo da especificidade.

Deve-se utilizar o roteiro da entrevista para auxiliar na elaboração dessas subcategorias. Quanto mais profunda e específica for a UCE, de menor nível será a subcategoria até o nível terciário. Observar que, algumas vezes, em vez de ser uma subcategoria de nível inferior, pode ser mais lógico gerar novas subcategorias primárias.

É mais fácil fazer esse procedimento individualmente em cada entrevista, isto é, mantendo-se arquivos separados de cada uma delas. Só depois da categorização individual, unem-se todas as entrevistas em um único arquivo, juntando-se as UCEs de mesma categoria e subcategoria, isto é, agrupando-as em uma estrutura hierárquica de assuntos que parte do geral para o específico.

Note, no entanto, que, para se fazer esse procedimento em arquivos separados, é necessário manter certa padronização nas subcategorias encontradas, de forma a não haver dificuldades e necessidade de revisão da classificação no momento de agrupar tudo em um só arquivo.

Para facilitar, é desejável que uma só pessoa realize a classificação e mantenha, em uma lista à parte, todas as subcategorias que for encontrando de forma organizada hierarquicamente. Essa lista deve ser consultada e atualizada a todo o momento, assim que forem sendo efetuadas as classificações de cada entrevista.

Na Figura 13, mostra-se a categorização da Classe 3 – Acervo eletrônico.

Categoria 3	Subcategoria primária	Subcategoria secundária	Subcategoria terciária	
Acervo eletrônico	Conhecimento da coleção	- Conhece		
		- Não conhece		
	Completeza da coleção			
	Desenvolvimento da coleção	- Resumos expandidos		
		- Outros periódicos		
	Artigos na mídia	- Tipos que a Unidade publica		
		- Tornar disponível	- Depende do tipo	
			- Totalmente a favor	
	- Seleção e Organização			
	Pôster	- Saber da existência não sei se ainda tem (pôster de trabalho)/[11] Eu acho que agora não se coloca mais (pôster) /[11] Eu acho que (pôster) não está sendo mais (colocado) /[11]		
- Tornar disponível		- A favor		
		- Contra Uma coisa que eu acho que não deveria colocar (...) mas pôster de trabalho/[11] Eu não acho que é muito bom colocar pôster/[11] muitas vezes, (...) não vou dizer que é todo mundo, mas muita gente coloca tópicos (nos pôsteres) /[11] (...) eu acho que (o pôster) não é uma coisa com muita serventia/[11] dependendo do que tem no pôster, (se você divulga) pode até te prejudicar na hora de publicar/[11] A pessoa olha lá e está disponível, por exemplo, na internet a informação (pôster)/[11] você pega (...) uma informação (pôster) que não é nem uma revista, nem nada/[11] É só um pôster que está na internet/[11] Eu acho que (...) (pôster) não deveria ser colocado/[11]		
- Duplicidade				
- Qualidade				

Figura 13. Categorização da classe 3 – acervo eletrônico.

Foi nessa classe, nas células de cor azul, onde se encaixaram as UCEs mostradas na Figura 12. De forma didática, evitando-se que a figura ficasse muito grande, suprimiram-se as outras UCEs, tanto aquelas de mesma classificação como as de outras subcategorias.

Elaboração da tabela de categorização de UCEs

Depois de se elaborar uma estrutura semelhante à Figura 13 para todas as classes, é necessário elaborar uma tabela resumida que conterá essa estrutura, porém com apenas duas a três UCEs em cada subcategoria existente. Essa tabela conterá, também, a frequência e porcentagens de UCEs em cada nível hierárquico, isto é, desde o nível hierárquico mais específico em que ela foi classificada, passando por todos os níveis hierárquicos, até o nível mais alto (total da classe) e total geral do *corpus*.

O motivo de não se colocar todas as UCEs nessa tabela é para evitar que ela fique muito pesada de informações e difícil para identificar as frequências e porcentagens de cada subcategoria e classe.

Na Figura 14, mostra-se exemplo da parte relativa à Classe 2 – Marketing do serviço oferecido. Cada subcategoria possui, no item “Exemplo das respostas”, duas UCEs pertencentes a ela.

A tabela (Figura 14) contendo as frequências e as porcentagens de UCEs será utilizada pelo pesquisador para a interpretação dos resultados. Note que a tabela é preparada subdividindo-se, hierarquicamente, primeiro as classes e, em seguida, as subcategorias primárias, secundárias e terciárias. O título da classe (categoria principal) encabeça a tabela. No nosso exemplo, é a Classe 2 – Marketing do serviço oferecido. Aparecem na primeira coluna três subcategorias primárias: Identificação da falta de divulgação; Formas de divulgação e Treinamento. O total e porcentagens de UCEs de cada uma dessas subcategorias são indicadas nas linhas de cor azul mais escuro. Nessa linha e nas colunas correspondentes aos subtotais, mostram-se, também, as porcentagens com relação à categoria e ao total geral do *corpus*.

Categoria primária II – Marketing do serviço oferecido							
Subcategoria 1	Subcategoria 2	Subcategoria 3	UCE (Exemplos de Respostas)	f	% (Subtotal)	% Categoria	% Total
Identificação da falta de divulgação	-	-	Eu não estou usando corretamente (o acesso à biblioteca eletrônica)/[2]	17	100%		
			Então eu estou entrando (...) (no site errado)/[2]				
			(...) eu caio em publicações, então minha dificuldade é com esse link (de publicações)/[2]				
Total da Subcategoria 1				17	100%	54,84%	2,61%
Formas de divulgação	Email	-	Foi muito interessante vocês terem mandado aqueles emails informativos sobre a utilização (da biblioteca eletrônica)/[2]	2	20%		
			estou gostando de ver a biblioteca (eletrônica), com os emails está mais dinâmica/[7]				
	Palestra	-	(...) (marcar a palestra em) uma semana que fosse mais tranquila/[2]	7	70%		
			Seria interessante (uma palestra para divulgar a biblioteca eletrônica)/[2]				
Dicas	-	Ali no link mesmo da biblioteca, (incluir) as dicas (de consulta) é importante também/[2]	1	10%			
Total da Subcategoria 1				10	100%	32,26%	1,54%
Treinamento	-	-	Se derem um treinamento (da biblioteca eletrônica) no auditório para todos (uma vez por ano) (...) já é suficiente/[7]	4	100%		
			(não precisa de treinamento)/[11]				
Total da Subcategoria 1				4	100%	12,90%	0,61%
Total da Categoria II				31	100%	100%	4,77%
Total Geral				650	100%	100%	100%

Figura 14. Frequência e porcentagem de UCEs por subcategorias.

A primeira subcategoria não possui subdivisões. Ela tem 17 UCEs que representam 100% da subcategoria. O assunto dessa subcategoria, conforme indicado pelo seu nome, é "Identificação da falta de divulgação". Sua importância na classe pode ser verificada pela porcentagem em relação ao total da categoria 2 (Marketing do serviço oferecido), cujo valor é de 54,84%, bastante significativo. Ela foi calculada multiplicando-se o total de UCEs da subcategoria (17) por 100 e dividindo-se esse total pelo total de UCEs da classe (31), isto é, $17 \times 100/31 \approx 54,84\%$.

A importância dessa subcategoria em relação ao total geral do corpus é calculada dividindo-se pelo total de UCEs do corpus, que em nosso caso é 650: $17 \times 100/650 \approx 2,61\%$, porcentagem bem pequena.

Já a subcategoria 2 (Formas de divulgação) tem o total de 10 UCEs, 32,26% da categoria 2 (Marketing do serviço oferecido), isto é, $10 \times 100/31 \approx 32,26\%$. Essa subcategoria possui 1,54% do total do corpus, isto é, $10 \times 100/650 \approx 1,54\%$. Ela se subdivide em subcategorias secundárias: Email com 2 UCEs, 20% da subcategoria ($2 \times 100/10$); Palestra com 7 UCEs, 70% da subcategoria ($7 \times 100/10$); e Dicas com 1 UCE, 10% da subcategoria ($10 \times 100/10$).

Em seguida, a subcategoria 3 (Treinamento) não possui subdivisões. Ela tem somente 4 UCEs que representam 100% da subcategoria, 12,90% da categoria ($4 \times 100/31$) e 0,61% do total geral ($4 \times 100/650$).

A penúltima linha da tabela, Total da categoria II, indica a frequência (31) e porcentagem de UCEs da subcategoria primária (100%), isto é, 20% + 70% + 10%. Indica, também, na coluna relativa ao subtotal da categoria, a porcentagem de 100%, calculada somando-se todos os subtotais dessa coluna que são relativos à classe em questão ($54,84\% + 32,26\% + 12,90\% = 100\%$). Indica, ainda, na coluna correspondente ao total geral, a porcentagem de 4,77%, calculado somando-se todas as porcentagens dessa coluna pertencentes à classe em questão, isto é, $2,61 + 1,54 + 0,61 \approx 4,77\%$, cujo valor representa a pequena força dessa classe com relação ao corpus total.

Já a última linha, Total Geral, é o fechamento e deve totalizar 100% nas três últimas colunas da tabela. Indica o total de UCEs do corpus (650), calculado somando-se os totais de UCEs de cada classe. Indica, também, a porcentagem de todas as UCEs (100%), calculando-se a soma das porcentagens de cada classe e dividindo-se o resultado pelo número de classes; e a porcentagem em relação a todo corpus, também (100%), somando-se a porcentagem total de todas as categorias. Esses cálculos não estão detalhados porque o exemplo mostrado limita-se à categoria II.

No exemplo mostrado, não existem casos de subcategorias terciárias. Porém, quando for necessário indicá-las, o processo é bastante simples. A terceira coluna (Subcategoria 3) é utilizada para indicar o seu nome. As demais colunas são preenchidas com os dados relativos a elas.

Pode, também, acontecer de haver mais de uma subcategoria secundária dentro de uma subcategoria primária. Nesse caso, mais um cálculo é efetuado, além do subtotal da subcategoria primária. Calcula-se a porcentagem de UCEs de cada subcategoria pertencente à mesma subcategoria secundária. Esse cálculo é colocado na linha de subtotal, semelhantemente à linha de subtotal da subcategoria primária.

Na Figura 15, mostra-se parte da Categoria primária I – Experiência do usuário na consulta do *site* e toda a subcategoria primária, Frequência de consulta, que se subdivide nas subcategorias secundárias: Periodicidade e Falta de tempo. Periodicidade, por sua vez, subdivide-se em 4 subcategorias terciárias: Quinzenal, Semanal, Esporádica e Não sabe.

Como existem duas subcategorias secundárias dentro de uma mesma subcategoria primária, coloca-se uma linha de subtotal entre elas.

Ao final da subcategoria primária, apresentam-se os totais somando-se os subtotais das categorias secundárias: frequência de UCEs ($36 + 13 = 39$); porcentagem de UCEs ($100\% + 100\%/2$); porcentagem da categoria ($8,26\% + 2,98\% = 11,24\%$); e porcentagem total do corpus ($5,54\% + 2,00\% = 7,54\%$).

Categoria primária I – Experiência do usuário na consulta do site							
Subcategoria 1	Subcategoria 2	Subcategoria 3	UCE (Exemplos de Respostas)	f	% (Subtotal)	% Categoria	% Total
Frequência de consulta	Periodicidade	Semanal	(...)(a frequência de uso da biblioteca eletrônica é semanal)/[2]	3	8,33%		
			Eu dou uma olhada (na biblioteca eletrônica semanalmente)/[7]				
		Quinzenal	(a frequência de consulta) acho que é meio que (a) cada 15 dias/[12]	2	5,55%		
			(...) (a frequência é quinzenal) por aí/[12]				
		Esporádica	Eu uso (...) (a Biblioteca eletrônica) com uma periodicidade esporádica/[3]	29	80,55%		
			(A consulta) é esporádica/[14]				
			(...)(a frequência de uso da biblioteca eletrônica) varia muito/[1]				
		Não sabe	(a frequência de consulta) vem de acordo com a (...) minha necessidade/[12]	2	5,55%		
			(A consulta da) (...) biblioteca eletrônica não tem questão de frequência/[12]				
				Não sei nem com que frequência (consulta)/[14]			
Subtotal				36	100%	8,26%	5,54%
Falta de tempo			(...) na verdade, a gente tem tido muito pouco tempo para pesquisar/[2]	13	100%		
			(...) A última coisa que a gente tem feito é estudar/[2]				
			(...) sentar numa biblioteca (...) é a última coisa que a gente tem feito/[2]				
Subtotal				13	100%	2,98%	2%
Total da subcategoria 1				49	100%	11,24%	7,54%

Figura 15. Exemplo de tabela com subcategorias terciárias.

Interpretação e redação dos resultados

Para interpretar os resultados, é necessário consultar o arquivo de categorização das classes e a tabela de frequências e porcentagens (Figura 14).

A análise é efetuada iniciando-se pela Categoria primária 1, seguindo por suas subcategorias primárias, secundárias e terciárias. A análise inclui descrição do assunto abordado pela categoria, subcategorias e considerações sobre o seu peso em relação ao discurso total e da categoria respectivamente.

É interessante apresentar nessa descrição uma figura contendo a estrutura hierárquica da categoria, incluindo os nomes das subcategorias e sua hierarquia, semelhantemente à Figura 13, porém, de forma mais simplificada, sem indicar as UCEs. Por meio dessa figura, o leitor tem uma visão rápida e geral da categoria.

Para se descrever o assunto do discurso tratado nas categorias e subcategorias, é necessário examinar as UCEs indicadas no arquivo de categorização. Ele contém todas as UCEs do discurso, pois o nome da categoria ou subcategoria não esclarece totalmente o assunto abordado, e a tabela de frequências lista apenas duas UCEs a título de exemplo.

Já a importância da categoria e subcategoria é verificada analisando-se sua respectiva frequência e porcentagem em relação ao todo (quando a análise é de uma categoria) ou à categoria (quando a análise é de uma subcategoria).

Cabe lembrar, porém, que os resultados de frequência e porcentagem referem-se às UCEs e não ao número de indivíduos. Essas UCEs podem ter sido proferidas por um só, dois ou mais indivíduos. O número de indivíduos não é considerado nessa análise.

Terminada a descrição da categoria 1, com todas suas subcategorias, é importante se elaborar uma conclusão geral dos resultados obtidos, ressaltando-se os pontos mais importantes.

As demais categorias são analisadas em sequência. No Anexo II, observa-se a descrição da categoria 3 Acervo eletrônico.

Análise das relações existentes entre as classes

No passo três, realiza-se a análise das relações existentes entre as classes, tomando-se como base o dendograma apresentado pelo software Alceste: apresentação das partições e da diferenciação do corpus, decorrentes do processo da Classificação Descendente Hierárquica (CDH).

É importante ressaltar, nessa análise, o total de classes encontradas e quais as partições existentes. Em seguida, cada classe deve ser detalhada com as seguintes informações: o título atribuído com base no tema do discurso da classe e seu total e percentual de UCEs; as variáveis descritivas relativas ao perfil que são significativas na classe; e as unidades lexicais predominantes na classe, com sua frequência (f) e qui-quadrado (χ^2).

Apresenta-se, na Figura 16, a descrição da classe 1 do dendograma apresentado na Figura 4.

A Classe I intitulada Experiência do usuário na consulta do site foi formada por 250 UCEs lexicais, que corresponde a 42,1% do corpus. O discurso dessa classe é marcado principalmente pelos respondentes caracterizados pelas variáveis descritivas:

Idade: menos de 30 anos (44 u.c.e.; $\chi^2 = 7$)

Sexo: feminino (129 u.c.e.; $\chi^2 = 5$)

Essa classe foi formada pela opinião dos entrevistados (mulheres com menos de 30 anos) a respeito dos recursos de consulta da Biblioteca eletrônica (veja, no último quadro do dendograma, a relação das raízes lexicais que formaram a classe). Salientamos os termos: tipo_de_public+ (tipo de publicação); título; autor+; consulta_simples, assunto; consulta_avan+ (consulta avançada); árvore_hiperbol+ (árvore hiperbólica); estrutura_de_a+ (estrutura de assunto), que indicam os recursos de consulta e as palavras: consult+; conhec+; dificuldade; fácil+; lembr+; prefiro; otim+; e gost+ como a opinião dos respondentes sobre elas.

Figura 16. Análise da Classe 1 do Dendograma.

Essa análise fecha o estudo do usuário interno. No tópico “Usuários Externos”, relata-se a metodologia utilizada para o estudo de usuários externos.

Usuários Externos

Instrumento de coleta de dados

O estudo de usuários externos foi realizado por meio de um questionário aplicado na página da Biblioteca Eletrônica por um período de seis meses (20 de dezembro de 2010 a 3 de junho de 2011). Foram coletadas informações sobre o perfil do usuário: características sociais e demográficas, tais como, sexo, idade, escolaridade, profissão, ocupação, instituição, cidade, estado, país e informações específicas sobre o site da Biblioteca Eletrônica:

- a) Avaliação da *homepage*: qual a facilidade, a capacidade de recuperação e a rapidez da busca de informação, terminologia e leiaute.
- b) Avaliação do conteúdo: assuntos de interesse.
- c) Impacto do serviço: frequência e objetivo de consulta.

As questões foram elaboradas em linguagem simples e objetiva, evitando-se o uso de terminologia inacessível à população em geral. O questionário encontra-se no Anexo III.

O pré-teste foi realizado aplicando-se o questionário por uma semana com a finalidade de averiguar complexidade das questões; ambiguidades; linguagem difícil; ordem das questões e tempo de duração.

Após identificação e correção dos problemas, foi efetuada a coleta pelo período especificado.

Escolha dos sujeitos

Visando aprimorar os resultados obtidos em 2009, quando foi realizada uma avaliação preliminar da Biblioteca Eletrônica, decidiu-se escolher os

respondentes de forma que os resultados fossem mais proveitosos. O objetivo foi selecionar aqueles usuários que já conheciam o site e que podiam opinar melhor.

Dessa maneira, assim que o usuário se cadastrava preenchendo as informações biodemográficas, era indagado sobre o número de vezes que já havia consultado a Biblioteca Eletrônica. Somente para aqueles usuários que já a haviam consultado por duas outras vezes foram aplicadas questões fechadas e abertas de avaliação da Biblioteca Eletrônica.

Análise dos resultados

Colheram-se 926 respostas do cadastro (média de 5,57 de usuários/dia). Desse número, 303 (média de 1,83) também responderam ao questionário. Significa que uma porcentagem alta, 32,85%, são usuários frequentes.

Os resultados foram analisados utilizando-se o pacote estatístico Statistical Package for the Social Sciences – SPSS, versão 18 e distribuídos em duas vertentes: (a) perfil biodemográfico; e (b) avaliação do site.

Importação e tratamento dos dados

Os dados provenientes da base de dados Postgres foram reformatados para Excel e, lá, sofreram uma primeira transformação. As respostas de perguntas abertas, como instituição, assunto pesquisado, opinião sobre leiaute, opinião sobre terminologia e sugestões, foram agrupadas por afinidades, evitando um leque muito grande de opções, e depois recodificadas em novas colunas, usando-se códigos padronizados. Criaram-se, portanto, tabelas de profissão e motivo da busca (que incluiu as profissões e os motivos das tabelas originais mais aqueles especificados pelo sujeito), assuntos, instituição, opinião sobre leiaute, opinião sobre terminologia e sugestões. Esse processo foi realizado com muita cautela para não descaracterizar os dados originais.

As instituições, depois de padronizadas, foram classificadas em 20 categorias, conforme Tabela 3.

Tabela 3. Tipo de instituição.

Código	Tipo de instituição
1	Extensão rural
2	Universidades e Escolas superiores
3	Colégios e outras instituições de ensino (exceto ensino superior)
4	Governo federal
5	Governo estadual ou municipal
6	Agronegócio (fazendas)
7	Outras empresas particulares
8	Particular (pessoa física)
9	Em branco
10	Associações, Fundações, Institutos e Ongs
11	Cooperativas
12	Bancos e assemelhados
13	Poder legislativo
14	Poder judiciário
15	Embrapa
16	Colégios agrícolas
17	Sindicatos
18	Outros ou desconhecido
19	Imprensa e Editoras
20	Escolas técnicas

Já os assuntos pesquisados foram classificados em temas e palavras-chaves mais específicas dentro de cada tema. Na Tabela 4, mostram-se os temas principais e, na Tabela 5, as palavras-chaves de Produção e Ciência animal.

Tabela 4. Temas principais.

Código	Tema
01	Produção e Ciência vegetal
02	Produção e Ciência animal
03	Ciência e tecnologia alimentar e industrialização de outros produtos agropecuários
04	Recursos naturais e meio ambiente
05	Agroenergia
06	Biologia geral

Continua...

Tabela 4. Continuação.

Código	Tema
07	Geografia e Geoprocessamento
08	Economia agrícola
09	Desenvolvimento rural
10	Implementos e engenharia rural
11	Química geral
12	Inválido
13	Banco de germoplasma
14	Outros (Administração, Tecnologia da informação, Disseminação Seletiva da Informação, Gestão e informações institucionais, História, Legislação, Política, Projeto Minibibliotecas e Atualidades).

Tabela 5. Palavras-chaves de Produção e Ciência animal.

Código	Assunto
16	Alimentação animal
23	Aquicultura
27	Avicultura
37	Biotechnology animal
41	Bovinocultura
42	Bovinocultura de corte
43	Bovinocultura de leite
54	Caprinocultura
59	Carrapatos
75	Confinamento
81	Controle da mosca dos chifres
83	Controle zootécnico
102	Equinocultura
163	Manejo animal
196	Nutrição animal
201	Ovinocultura
207	Pecuária
221	Produção animal
222	Produção de leite
223	Produção de leite orgânico
225	Produção de ovos

Continua...

Tabela 5. Continuação.

Código	Assunto
230	Qualidade do leite
247	Reprodução animal
263	Suínocultura
264	Suplemento alimentar
273	Veterinária
277	Zoologia

Os dados assim preparados foram importados no SPSS. Durante o procedimento de importação, especificaram-se em *Range* as colunas e linhas inicial e final da planilha do Excel de onde os dados foram importados. Ex.: A4-IA264. Esse processo permitiu indicar quais registros e colunas deveriam ser importados, quando somente uma parte da planilha será importada. Porém, mesmo que seja a planilha inteira a ser importada, é interessante especificar essas informações para evitar erros.

Após a importação, acessou-se a tela *Variable view* e se definiram adequadamente as variáveis: nome, tipo (numérico, *string*, data, etc.), comprimento, etiqueta (nome por extenso da variável), valores que pode assumir, alinhamento desejado (esquerda, direita ou centro) e medida (nominal, ordinal ou escalar).

Em seguida, os dados foram verificados na tela *Data view*, a procura de *outliers*, isto é, dados discrepantes e registros com erro de gravação. Quando a quantidade de dados é grande é mais fácil examinar esses problemas por meio da emissão prévia das frequências e porcentagens de todas as variáveis. Caso tenha havido valores que possam atrapalhar a análise, deve-se decidir sobre permanecer com os dados na amostra, substituí-los pelos valores da média ou, ainda, eliminá-los.

Todas as variáveis de tipo *string* utilizadas para cálculo de frequência e porcentagens, tais como país, região, sexo, profissão, foram tratadas previamente, isto é, convertidas para código numérico.

Existem procedimentos no SPSS bastante úteis para realizar esse trabalho no menu *Transform*, tais como, *Recode into different variables*; *Automatic recode*; *Compute variable*. Existe também, no referido menu, tratamento para os *missing* (valores em branco).

Em seguida, foram estudadas formas para condensar as respostas para as variáveis de largo espectro como idade, estado, país estrangeiro. Foram, então, criadas outras variáveis utilizando-se o menu *Transform*. Foram criadas regiões geográficas, países estrangeiros, a partir das variáveis Estado e País, respectivamente, e faixas para Idade.

Estatística descritiva

As primeiras análises foram de Estatística descritiva com emissão das tabelas de frequências e porcentagens para todas as variáveis, cálculo da média, moda e desvio-padrão (MARTINEZ; FERREIRA, 2010).

A partir dos resultados obtidos, estudaram-se quais variáveis deveriam ser cruzadas, a fim de se identificar grupos de perfis. Realizaram-se os seguintes cruzamentos: Instituição x Região brasileira; Instituição x Exterior; Tema x Região brasileira; Palavras-chaves x Região brasileira.

Além disso, verificou-se que, apesar de existir um perfil de usuário predominante, parte significativa dos resultados apresentou grande variação em idade, profissão e instituição. Dessa forma, considerou-se importante estratificar as análises por grupos e verificar as diferenças de opinião entre eles.

Vantagens da metodologia

A estatística quantitativa permite a análise rápida e precisa de grande quantidade de dados. Ela permite que se obtenha um panorama geral do fenômeno estudado. E, ainda, fornece direcionamentos para se identificar pontos obscuros, dúvidas ou problemas que serão analisados em estudos posteriores mais detalhados e profundos, utilizando métodos qualitativos (NEVES, 1996).

Já a análise qualitativa é aplicada a uma amostra pequena porque demanda muito esforço para coleta (efetuada por meio de entrevistas)

e, principalmente, para análise dos dados que é manual. Ela é mais adequada para ser realizada a partir dos resultados da análise quantitativa, direcionando a indagação para os pontos sugeridos e, para aqueles em que se deseja detalhar.

A análise qualitativa, ao contrário, pode ser também o primeiro estudo realizado. Um estudo posterior que utilizará métodos quantitativos pode confirmar e generalizar os resultados já obtidos (SERAPIONI, 2000; NEVES, 1996).

A realização de uma ou outra depende muito do objeto e propósito de estudo. Algumas vezes é interessante iniciar com a análise mais abrangente (quantitativa) para definir as variáveis a serem estudadas e, então, aprofundar o estudo (análise qualitativa) (ROCHA et al., 2010; ROCHA et al., 2008).

A análise assim combinada é complementar (MINAYO; SANCHES, 1993). Apresenta como vantagens minimizar a subjetividade do método qualitativo, e, em contrapartida, aproximar o pesquisador do objeto de estudo, respondendo, assim, às principais críticas das abordagens somente qualitativas ou quantitativas, agregando maior confiabilidade aos dados.

A realização do estudo de usuários por meio da metodologia mista ora descrita permitiu não somente se obter uma visão global de como se encontra o site na opinião de uma amostra de usuários internos e externos, mas também saber quais aspectos concretos e específicos que necessitam ser revistos e melhorados.

Conclusão

Descreveu-se neste trabalho uma metodologia mista, que combinou a análise de dados quantitativos e qualitativos, para realizar o estudo de usuários da Biblioteca Eletrônica da Embrapa Cerrados.

O estudo dividiu-se em duas vertentes: estudo dos usuários internos (empregados, bolsistas e estagiários) e estudo dos usuários da internet, cada um utilizando metodologias distintas para coleta e análise dos dados.

O estudo dos usuários internos foi realizado por meio de entrevista semiestruturada, gravada, transcrita e analisada por meio de método qualitativo. Utilizaram-se o software Alceste, versão 2009 para Windows, que realiza análise estatística de coocorrência de palavras do discurso das entrevistas e, em seguida, a análise de conteúdo de Bardin, efetuada em 50% da amostra, a qual possibilitou obter resultados mais detalhados e profundos.

O uso da metodologia conjugada Alceste e Bardin evitou obter resultados subjetivos, situação comum de acontecer com o uso apenas da análise de conteúdo de Bardin. Por ela ser totalmente manual pode, até certo ponto, ser influenciada pelo especialista que a realizou, causando resultados diferentes para a divisão do *corpus* em categorias, se realizada por diferentes especialistas.

Já com a utilização do software Alceste, a quantidade de categorias primárias é estabelecida estatisticamente, a partir da coocorrência de palavras e da análise do qui-quadrado, diminuindo essa subjetividade.

A partir dessa categorização inicial, realiza-se a análise de conteúdo, dividindo o discurso de cada categoria em subcategorias, quantas forem necessárias para detalhar todos os tópicos de assunto tratados.

Cabe ressaltar, no entanto, que foi efetuada uma modificação no método tradicional. No método de análise pelo software Alceste, preconiza-se: a não alteração na redação das entrevistas transcritas, apenas a correção dos sinais de pontuação e da grafia de palavras; e a eliminação de parte do discurso que não trata do objeto de análise.

Porém, na metodologia proposta neste trabalho e, efetivamente, testada na análise das entrevistas da Biblioteca Eletrônica, houve uma modificação dessa regra. As razões dessa alteração são: o português falado difere muito do escrito; muitas frases são inacabadas; o discurso é formado por palavras repetidas, gírias, neologismos e formas contraídas; além disso o receptor capta o discurso não só pelas palavras, mas também por gestos, expressões faciais, postura, que muitas vezes completam ou esclarecem o sentido da comunicação.

Por isso, o *corpus* submetido à análise automática foi adaptado, utilizando-se o mesmo procedimento daquele adotado na análise de conteúdo. Acredita-se que essa alteração proporcionou resultados melhores e mais eficientes.

Sugere-se, também, outra alteração que poderá ser testada em estudos futuros. A metodologia de análise de conteúdo não é voltada para medir quantitativamente a opinião em termos de indivíduos, mas ao peso daquela resposta com relação à categoria, subcategoria ou *corpus*. Esse peso pode ter sido dado por apenas um indivíduo, dois ou, sucessivamente, até por todos os indivíduos da amostra. Pois, o cálculo das frequências das UCEs não leva em consideração quantos indivíduos opinaram, isto é, a opinião a favor ou contra de determinado objeto pode ser determinada por apenas um indivíduo em contraposição ao grupo. Se, por exemplo, um só indivíduo gerou 40 UCEs a favor e 20 outros indivíduos foram contra e geraram no total 30 UCEs, o peso maior da discussão é de apenas um indivíduo que provocou maior número de UCEs.

Seria, portanto, interessante que o investigador realizasse, também, a contabilização de frequência e porcentagens de UCEs por indivíduos. Seria um elemento a mais para análise que contemplaria a opinião do grupo e não apenas de indivíduos. O acréscimo desses novos cálculos orientaria o especialista na discussão dos resultados e conclusão do estudo.

Os usuários externos foram investigados por meio de um questionário postado na página da Biblioteca Eletrônica.

Os dados levantados por questões abertas foram transformados em respostas padronizadas e agrupados em faixas. Depois, juntamente com os outros oriundos de perguntas fechadas, foram analisados estatisticamente por meio do SPSS, gerando tabelas de frequências, porcentagens, média, moda e desvio-padrão.

Em seguida, analisando-se esses resultados, escolheram-se as variáveis que seriam interessantes realizar cruzamentos, a fim de se estabelecer perfis de grupos de usuários.

Os resultados dos estudos de usuários realizados com a utilização dessa metodologia forneceram informações úteis, esclarecedoras e ricas em detalhes que nos possibilitou melhorar o site da Biblioteca Eletrônica e traçar o perfil de seus principais usuários. Esses resultados dificilmente seriam alcançados com o uso apenas de métodos quantitativos.

Referências

- ALCESTE: un logiciel d'analyse de données textuelles. Outil d'aide à l'analyse automatique rapide et efficace. Toulouse: Image, [s.d.]
- BAPTISTA, S. G.; CUNHA, M. B. da. Estudo de usuários: visão global dos métodos de coleta de dados. **Perspectivas em Ciência da Informação**, v. 12, n. 2, p. 168-184, maio/ago. 2007.
- BARBETTA, P. A. **Estatística aplicada às ciências sociais**. 5. ed. Florianópolis: Editora da Universidade Federal de Santa Catarina, 2002.
- BARDIN, L. **Análise de conteúdo**. Lisboa: Edições 70, 2004.
- CARVALHO, N. **Publicidade**: a linguagem da sedução. São Paulo: Ática, 2003.
- COSTA, L. F. da; SILVA, A. C. P. da; RAMALHO, F. A. (Re)visitando os estudos de usuário: entre a "tradição" e o "alternativo". **Revista de Ciência da Informação**, v. 10, n. 4, agosto de 2009.
- COZBY, P. **Métodos de pesquisa em ciências do comportamento**. São Paulo: Atlas, 2003.
- FIGUEIREDO, N. M. de. **Estudos de uso e usuários da informação**. Brasília, DF: IBICT, 1994. 154 p.
- FONTANELLA, B. J. B.; RICAS, J.; TURATO, E. R. Amostragem por saturação em pesquisas qualitativas em saúde: contribuições teóricas. **Cadernos de Saúde Pública**, v. 24, n. 1, p. 17-27, jan. 2008.
- LEBART, L.; SALEM, A. **Statistique textuelle**. Paris, 1994.
- MARTINEZ, L. F.; FERREIRA, A. I. **Análise de dados com SPSS**: primeiros passos. 3. ed. Lisboa: Escolar Editora, 2010. 177 p.
- MINAYO, M. C. S. (Org.). **Pesquisa social**: teoria, método e criatividade. Petrópolis: Vozes, 1994.
- MINAYO, M. C. S.; SANCHES, O. Quantitativo-qualitativo: oposição ou complementaridade? **Cadernos de Saúde Pública**, v. 9, n. 3, p. 237-48, Jul./Set. 1993.

NASCIMENTO, A. R. A.; MENANDRO, P. R. M. Análise lexical e análise de conteúdo: uma proposta de utilização conjugada. **Estudos e Pesquisas em Psicologia**, v. 6, n. 2, p. 72-88, jul./dez. 2006.

NEVES, J. L. Pesquisa qualitativa: características, usos e possibilidades. **Caderno de Pesquisas em Administração**, v. 1, n. 3, 1996.

RICHARDSON, R. J.; PERES, J. A. de S.; WANDERLEY, J. C. V.; CORREIA, L. M.; PERES, M. de H. de M. **Pesquisa social: métodos e técnicas**. São Paulo: Atlas, 1999.

ROCHA, F. E. de C.; ALBUQUERQUE, F. J. B. de; COELHO, J. A. P. de M.; MARCELINO, M. Q. dos S. **Elaboração de Instrumento de Avaliação do Pronaf para Agricultores com Baixa Escolaridade**. Planaltina, DF: Embrapa Cerrados, 2008. 58 p. (Embrapa Cerrados. Documentos, 210).

ROCHA, F. E. de C.; MARCELINO, M. Q. dos S.; KISHI, S. M.; Malaquias, J. V. **Método de análise de conteúdo com categorização apriorística baseada na Teoria da Ação Planejada**: uma avaliação da adoção de práticas conservacionistas de recursos hídricos. Planaltina, DF: Embrapa Cerrados, 2011a. 142 p. (Embrapa Cerrados. Boletim de Pesquisa & Desenvolvimento, 305).

ROCHA, F. E. de C.; RAMOS, E. M. S.; KISHI, S. M.; MARCELINO, M. Q. dos S.; SOUSA, E. dos S. de; JÚLIO, L. de; TIMPONI, M. R. **Avaliação Qualitativa da Percepção de Extensionistas Rurais em Relação ao Uso e à Conservação de Recursos Hídricos**. Planaltina, DF: Embrapa Cerrados, 2010. 46 p. (Embrapa Cerrados. Boletim de Pesquisa & Desenvolvimento, 297).

ROCHA, F. E. de C.; MARCELINO, M. Q. dos S.; MARTINS, C. R.; KISHI, S. M.; SANTOS, L. P. **Avaliação qualitativa do uso da água na agricultura por meio da análise de conteúdo conjugada: modelo de Bardin e software Alceste**. Planaltina, DF: Embrapa Cerrados, 2011b. (Embrapa Cerrados. Boletim de Pesquisa & Desenvolvimento, 35).

SALVIATI, M. E.; CASTRO, R. L. de. **Biblioteca eletrônica da Embrapa Cerrados: modelo conceitual, desenvolvimento, implantação e avaliação**. Planaltina, Df: Embrapa Cerrados, 2009. 81 p. (Embrapa Cerrados. Documentos, 245).

SANZ CASADO, E. **Manual de estudios de usuarios**. Madrid: Pirámide, 1994.

SERAPIONI, M. Métodos qualitativos e quantitativos na pesquisa social em saúde: algumas estratégias para a integração. **Ciência & Saúde Coletiva**, v. 5, n. 1, p. 187-192, 2000.

SOFTWARE ALCESTE: analyse lexicale par context dún ensemble de segments de texte. Toulouse: Image, 2009.

SOUSA, E. dos S. de; RODRIGUES, M. A. dos S.; ROCHA, F. E. de C.; MARTINS, C. R. **Guia de utilização do software Alceste**: uma ferramenta de análise lexical aplicada à

interpretação de discursos de autores na agricultura. Planaltina, DF: Embrapa Cerrados, 2009. 37 p. (Embrapa Cerrados. Documentos, 275).

VALLS, V. M.; VERGUEIRO, W. de C. S. A gestão da qualidade em serviços de informação no Brasil: uma nova revisão de literatura, de 1997 a 2006. **Perspectivas em ciência da informação**, v. 11 n. 1, p. 118-137, jan./abr. 2006.

Proposed Methodology for the Study of Users of Embrapa

Abstract

This paper describes a methodology used for conducting the user study of Electronic Library of Embrapa Cerrados. The most recent user studies use both quantitative and qualitative methods. Quantitative research is characterized, both in the data collection and in their treatment by the use of statistical techniques. On the other hand, to study human behavior is used qualitative methods. They seek to clarify who are the users of information, what their needs are and how to seek and use information. Take into account also the aspects of interaction between users and technology involved in the information service. The methodology described performs the collection of quantitative and qualitative data using two different techniques for collecting and analyzing data. The collection of data from the internal public of the Company was conducted through interviews and submitted to automatic content analysis, using the Alceste software in conjunction with Bardin method. A second data collection was performed with the internet public, through a questionnaire posted on the site of the Electronic Library. This latter was analyzed statistically using the SPSS software. This paper discusses not only the theoretical aspects of the methodology, but introduces some real and practical examples at every stage of the process, such as sampling, data collection instrument, operationalization of the data and its analysis.

Index terms: user study, methodology, quantitative analysis, qualitative analysis, content analysis, electronic library.

Anexo I. Roteiro da entrevista semiestruturada

Estudo de Usuários das Bibliotecas Eletrônicas da Embrapa Cerrados, Embrapa Hortaliças, Embrapa Informação Científica e Tecnológica e Embrapa Recursos Genéticos e Biotecnologia

Uso da biblioteca eletrônica

Nº da entrevista: _____

Data: ____/____/____

Fase I – Autoaplicada focada no perfil dos respondentes

1. De qual centro da Embrapa o(a) Senhor(a) faz parte?

- () Cerrados
- () Hortaliças
- () Recursos Genéticos e Biotecnologia
- () Informação Tecnológica
- () Sede

2. Qual o seu vínculo com a Embrapa?

- () Pesquisador
- () Analista
- () Bolsista
- () Estagiário

3. Sexo:

- () Masculino
- () Feminino

4. Idade:

- Menos de 30 anos
- De 30 a 39 anos
- De 40 a 49 anos
- De 50 a 59 anos
- 60 anos ou mais

5. Escolaridade:

- Ensino superior
- Especialização
- Mestrado
- Doutorado
- Pós-doutorado

6. Anos de trabalho na Unidade

- Menos 2 anos
- Entre 2 e 5 anos
- Entre 6 e 10 anos
- Entre 10 e 15 anos
- Mais de 15 anos

7. Anos de trabalho fora da Embrapa

- Menos de 2 anos
- Entre 2 e 5 anos
- Entre 6 e 10 anos
- Entre 10 e 15 anos
- Mais de 15 anos

8. Frequência de busca de informação pela Internet

- diária
- semanal
- mensal
- esporádica

9. Quais os *sítes* mais úteis para busca de informação na Internet?

10. O(A) Senhor(a) conhece a Biblioteca eletrônica da sua Unidade (/Sede)?

() Não **(Encerrar a entrevista aqui)**

() Sim

11. O(A) Senhor(a) já utilizou a Biblioteca eletrônica da sua Unidade (/ Sede)?

() Não

() Sim **(Pular para a questão 13)**

12. Por que o(a) Senhor(a) não a utiliza?

13. Que tipo de divulgação o(a) Senhor(a) sugere para aumentar o número de usuários?

14. Há necessidade de treinamento para a utilização da Biblioteca eletrônica ?

() Não

() Sim.

Encerrar a entrevista aqui para aqueles que responderam Não na pergunta 11 (não utilizam a biblioteca eletrônica)

Fase II – entrevista focada no uso eletrônico da biblioteca

15. Qual a frequência de utilização?

() diária

() semanal

() mensal

() esporádica

16. Quais os tipos de consulta que o(a) Senhor(a) já utilizou? Autor, Título, Tipo de publicação, árvore hiperbólica, lista estruturada de assunto, consulta simples, consulta avançada. Comente cada uma das utilizadas sobre a rapidez, facilidade e eficiência da consulta.

Rapidez: São as páginas são carregadas rapidamente ?

Facilidade: O *site* é fácil de consultar ou há necessidade de ajuda e treinamento?

Eficiência: A informação desejada é obtida em tempo razoável ?

17. O(A) Senhor(a) conhece qual o tipo das publicações incluídas na Biblioteca eletrônica? Deveria incluir outros tipos ? Quais?

18. Existe algum comentário ou recomendação a ser feita com relação à Biblioteca eletrônica?

Anexo II. Descrição dos resultados da Categoria III Acervo eletrônico.

Categoria 3 – Acervo eletrônico

Essa categoria possui 183 UCEs semânticas equivalentes a 28,15% do total de UCEs do discurso de oito entrevistas. Trata sobre o acervo eletrônico que é disponibilizado pelo site da Biblioteca eletrônica. A maior parte do discurso está concentrada em Pôster com 97 UCEs (53%) da categoria e Artigos na mídia com 56 UCEs (30,60%) da categoria. Outras subcategorias primárias com menor representatividade estão: Conhecimento da coleção com 8 UCEs (4,40%) da categoria; Completeza da coleção com 9 UCEs (4,92%) da categoria; e Desenvolvimento da coleção com 13 UCEs (7,10%), Tabela 1.

Tabela 1. Sumarização da categoria 3 e suas respectivas subcategorias.

Categoria 3	Subcategoria primária	Subcategoria secundária	Subcategoria terciária
Acervo eletrônico	Conhecimento da coleção	- Conhece	
		- Não conhece	
	Completeza da coleção		
	Desenvolvimento da coleção	- Resumos expandidos	
		- Outros periódicos	
		- Tipos que a Unidade publica	
	Artigos na mídia	- Tornar disponível	- Depende do tipo
		- Seleção e Organização	- Totalmente a favor
		- Saber da existência	
	Pôster	- Tornar disponível	- A favor
- Duplicidade		- Contra	
- Qualidade			

Subcategoria primária Conhecimento da coleção

Essa subcategoria expressa o conhecimento ou não da coleção que forma o acervo digital da Biblioteca eletrônica: 6 UCES (75%) afirmaram conhecer e 2 UCES (25%) afirmaram não conhecer. Também nessa subcategoria nota-se a falta de divulgação do serviço.

Subcategoria primária Completeza da coleção

Nessa subcategoria, os entrevistados representados por 9 UCES (4,92%) disseram estar completo o conteúdo da Biblioteca eletrônica e não recomendaram o acréscimo de mais nenhum tipo de publicação.

Subcategoria primária Desenvolvimento da coleção

Em contrapartida à subcategoria anterior, 12 UCES (7,10%) da categoria expressaram desejo de se acrescentar outros tipos de publicação ao acervo: Resumos expandidos com 9 UCES (69,23%) da subcategoria e Outros periódicos com 4 UCES (30,76%). Apresentam-se detalhes a seguir.

- **Resumos expandidos 9 UCES (69,23%)**: expressam o desejo de que sejam disponibilizados os resumos expandidos de congressos, pois eles são tão completos quanto um artigo.

- **Outros periódicos 4 UCES (30,76%)**: expressam o desejo de se acrescentar artigos da revista Pesquisa Agropecuária Brasileira e Revista Brasileira de Ciência do Solo. Recomendam, também, que se verifiquem quais outros periódicos poderiam ser legalmente incluídos.

Subcategoria primária Artigos na mídia

Nessa subcategoria os entrevistados discutem a importância, necessidade e utilidade de se incluir no acervo da biblioteca eletrônica os artigos publicados na mídia. O discurso total possui 56 UCES da categoria e foi classificado nas seguintes subcategorias: Tipos que

a Unidade publica 5 UCEs; Tornar disponível 40 UCEs; e Seleção e Organização 11 UCEs. Apresentam-se a seguir essas subcategorias com mais detalhes.

- Subcategoria secundária Tornar disponível

Discute-se nessa subcategoria se é desejável ou não tornar disponível na Biblioteca eletrônica os artigos na mídia (40 UCEs da subcategoria). As considerações a esse respeito foram classificadas em dois grupos: Depende do tipo com 6 UCEs; e Totalmente a favor 34 UCEs, conforme detalhes a seguir.

- **Depende do tipo** 6 UCEs (10,71%): discorre sobre os tipos de artigos da mídia existentes: notícias e matérias jornalísticas. Segundo eles, as notícias são efêmeras e, portanto, não há tanto interesse em mantê-las guardadas no acervo. Já as matérias jornalísticas são informações importantes, semelhantes a artigos e deveriam ser disponibilizadas.

- **Totalmente a favor** 34 UCEs (60,71%): em contrapartida ao discurso do item anterior, nesse item, os entrevistados afirmam a validade, importância e utilidade de se disponibilizar os artigos na mídia sem restrições, lembrando inclusive a importância histórica de se guardar esse material.

- Subcategoria secundária Seleção e organização

É discutida a necessidade de se organizar e armazenar os artigos da mídia e de preferência fazer uma separação entre as notícias e as matérias jornalísticas. Esse arquivo serviria, também, para se conhecer o que foi publicado e com qual qualidade. Contém 11 UCEs (19,64%) do total da subcategoria.

- Subcategoria secundária Tipos que a Unidade publica

Discorre sobre a diferença existente entre notícia e matéria jornalística. Contém 5 UCEs (8,93%) da subcategoria.

Anexo III. Questionário aplicado aos usuários externos.

Nome completo	<input type="text"/>	* Sexo:	Idade:
E-mail	<input type="text"/>	*	
Senha	<input type="text"/>	* Confirmação	<input type="text"/>
Profissão	<input type="text" value="Agrônomo"/>	* Em caso de Outras, especificar:	
Instituição	<input type="text"/>		
	Em caso de autônomo ou particular, favor explicitar essa condição		
Endereço correspondência	<input type="text"/>	*	
Cidade	<input type="text"/>	*	
Estado	<input type="text" value="DF"/>	*	
CEP	<input type="text"/>		
País	<input type="text" value="Brasil"/>	*	
Quantas vezes você já procurou informação em nossa Biblioteca eletrônica?			
Nenhuma vez			
Uma vez			
Algumas vezes			
Muitas vezes			

1. Com qual frequência você realiza busca de informação técnica na Internet ?

Diária

Semanal

Mensal

Ocasional

2. Qual o motivo desta busca (de hoje) na Biblioteca Eletrônica? Marque quantas opções forem necessárias.

a) Elaboração de pesquisas científicas

a) Elaboração de relatórios, planos e programas

a) Atualização do conhecimento

b) Elaboração de cursos, palestras, dias de campo e outras atividades relacionadas

a) Atendimento de consultas

b) Agronegócio

- Outro, especificar:

3. Você sabe que assuntos existem na Biblioteca eletrônica ?

Sim

Não

4. Qual(is) assunto(s) você está procurando?

a) _____

b) _____

c) _____

d) _____

e) _____

5. Qual o grau de facilidade do mecanismo de busca?

muito fácil

fácil

satisfatório

difícil

muito difícil

6. Como você avalia o tempo de resposta (tempo de carga das páginas)?

péssimo

satisfatório

bom

ótimo

7. Qual a sua opinião sobre as instruções de consulta?

Não consultei as instruções

péssima

satisfatória

boa

ótima

8. Qual o grau de facilidade de navegação?

muito fácil

fácil

satisfatório

difícil

muito difícil

9. Qual a sua opinião sobre o leiaute (aparência) das páginas?

10. Qual a sua opinião sobre a terminologia (palavras-chaves) utilizadas na estrutura de assuntos?

11. Que sugestões você daria para melhorar o site?

Embrapa

Cerrados

Ministério da
Agricultura, Pecuária
e Abastecimento

GOVERNO FEDERAL
BRASIL
PAIS RICO É PAIS SEM POBREZA

CGPE 10848