



UNIVERSIDADE DE BRASÍLIA
FACULDADE DE CIÊNCIA DA INFORMAÇÃO
GRADUAÇÃO EM BIBLIOTECONOMIA

JAQUELINE RODRIGUES DE JESUS

MINERAÇÃO DE DADOS: uma possibilidade de ferramenta para a tomada de decisão na
representação temática da informação

JAQUELINE RODRIGUES DE JESUS

MINERAÇÃO DE DADOS: uma possibilidade de ferramenta para a tomada de decisão na representação temática da informação

Monografia apresentada como pré-requisito para obtenção do título de Bacharel em Biblioteconomia pela Faculdade de Ciência da Informação (FCI) da Universidade de Brasília (UnB).

Orientador: Professor Mestre Márcio Bezerra da Silva

Dados Internacionais de Catalogação na Publicação (CIP)

J58m Jesus, Jaqueline Rodrigues de
Mineração de dados : uma possibilidade de ferramenta
para a tomada de decisão na representação temática da
informação / Jaqueline Rodrigues de Jesus. – Brasília, 2015.
88 f. : il.

Orientador: Professor Mestre Márcio Bezerra da Silva
Monografia (Graduação) – Universidade de Brasília,
Faculdade de Ciência da Informação (FCI), 2015.

1.Representação temática da informação. 2.Folksonomia.
3.Mineração de dados. I. Título.

CDU 002:004



Título: Mineração de dados: uma possibilidade de ferramenta para tomada de decisão na representação temática da informação .

Aluna: Jaqueline Rodrigues de Jesus.

Monografia apresentada à Faculdade de Ciência da Informação da Universidade de Brasília, como parte dos requisitos para obtenção do grau de Bacharel em Biblioteconomia.

Brasília, 14 de fevereiro de 2015.

Márcio Bezerra da Silva - Orientador
Professor da Faculdade de Ciência da Informação (UnB)
Mestre em Ciência da Informação

Dulce Maria Baptista – Membro interno
Doutora em Ciência da Informação

Ernani Rufino dos Santos Junior – Membro externo
Mestre em Ciência da Informação

*Dedico, a mais bela de todas as guerreiras,
minha Mãe Maria Estela.*

AGRADECIMENTOS

Em primeiro lugar, sempre a Deus.

Como me lembrar de todos os que me ajudaram a percorrer esta longa estrada?! Somente Deus para enriquecê-los de graças por me ajudarem nesta caminhada edificadora.

Começo a agradecer a aquela que me abasteceu de força e coragem quando a paciência e o amor pareciam acabar: Mãe, obrigada por tudo, simplesmente tudo. Palavras são incapazes de descrever meus agradecimentos a senhora. Ao meu pai, irmãos e sobrinhos, que fazem dias nublados aparecerem o sol.

Ao meu professor e orientador Márcio Bezerra da Silva, que me deu valiosas oportunidades de aprendizagem e pela imensa paciência por exercer seu trabalho, ensinando-me a cada dificuldade.

Agradeço aos outros principais docentes da Faculdade de Ciência da Informação (FCI), Fernanda Moreno, Murilo Bastos Cunha, Mamede Lima-Marques, Rita de Cássia Caribé, que ao longo dos quatros anos me ensinaram tudo o que podiam e achavam necessário para meu crescimento acadêmico e profissional. E ao querido secretário da FCI, Reginaldo Oliveira, que sempre nos ajudava nas burocracias acadêmicas.

Não poderia deixar de agradecer a aqueles que ouviram todas as minhas lamúrias, chatices, loucuras, sonhos e alegrias, meus amigos acadêmicos, que com certeza os levarei para a vida: Fernanda Maciel, Fernanda Diogo, Priscila Rodrigues, Inês Iacira, Elton Mártires, Maria Helena, Hélio Delfino, Daniel Rodrigues e Bruno Cunha.

Agradeço aos meus supervisores e colegas do Centro de Pesquisa em Arquitetura da Informação (CPAI) e Tribunal Regional Federal (TRF), onde realizei os meus estágios, permitindo-me aprimorar a ideia deste trabalho e aplicar aquilo que apreendi.

Termino agradecendo aqueles preciosos amigos, que mesmo indiretamente me incentivavam na vida acadêmica todos os dias, que fosse por um sorriso, ou um jeito de me fazer sorrir. Um abraço aos meus amigos da vida: Karine Macedo, Jéssica Patriarca, Valéria Oliveira, Ygor Pereira, Felipe de Jesus, Thiago Rodrigues, Nathália Santos e Ana Beatriz Santos. Todos são peças essenciais para me manter em um bom funcionamento psicológico.

Obrigada a todos!

*“Não basta saber, é preciso também aplicar,
não basta querer, é preciso também agir”.*
(Goethe)

RESUMO

Apresenta o uso da mineração de dados na representação temática da informação. Discute-se na fundamentação teórica os elementos que constituem a sociedade entre dado, informação, conhecimento, desenvolvimento e sabedoria; a representação temática da informação, sendo a indexação tradicional e folksonomia; e a mineração de dados. Objetiva-se analisar a mineração de dados como ferramenta para a tomada de decisão na representação temática da informação. Constitui-se de um estudo exploratório, de natureza aplicada, que utiliza a técnica de pesquisa bibliográfica, com abordagem de coleta de dados quantitativa e qualitativa, e que adota a ferramenta de mineração de texto *Kitconc* 3.0 como instrumento de coleta de dados para analisar os comentários mais recentes e definidos como úteis pelos usuários em oito livros disponibilizados no ambiente de comércio eletrônico *Amazon*, em seguida definir as palavras-chaves de cada obra e confrontá-las com os termos indexadores dos mesmos livros no catálogo da Biblioteca Central da Universidade de Brasília (BCE/UnB). Apresenta como resultados da pesquisa que a mineração de texto contribuiu de forma significativa e fidedigna para a descoberta de padrões para a representação temática da informação a partir dos termos que se apresentaram com frequência e possuíam concordância semântica no texto, tornando-os posteriormente candidatos às palavras-chave que permitirão definir os assuntos das obras. Resulta também, por outro lado, que a mineração de texto não identificou padrões suficientes em duas obras, impossibilitando realizar o confronto entre os dois ambientes. Conclui-se que os comentários obtidos na *web* 2.0 apresentam o conhecimento dos usuários que usufruíram da obra e explicitaram sua sabedoria em postagens, aqui transformadas em palavras-chave, mostrando que os mesmos podem participar efetivamente na representação temática da informação e contribuir posteriormente no sucesso dos processos de busca, nos permitindo inferir que para uma possível representação fidedigna do conteúdo da obra e melhor recuperação, a indexação social pode subsidiar/complementar a indexação tradicional, ou seja, as *tags* dos usuários, analisadas pela mineração, podem ser comparadas aos descritores definidos pelos bibliotecários para a tomada de decisão quanto aos termos representativos das obras.

Palavras-chave: Conhecimento. Sabedoria. Representação temática da informação. Folksonomia. Mineração de dados. Mineração de texto.

ABSTRACT

Presents the use of data mining in the thematic representation of information. It discusses in the theoretical basis the elements of society among data, information, knowledge, wisdom and development; the thematic representation of information, with the traditional and folksonomy indexing; and data mining. The objective is to analyze the data mining as a tool for decision making in the thematic representation of information. It consists of an exploratory study, of an applied nature, using the technical literature, with collection of quantitative and qualitative data approach, and adopting the Kitconc 3.0 text mining tool such as data collection instrument to analyze the most recent comments and defined as helpful by users in eight books available on Amazon e-commerce environment, then define the keywords of each work and compare them with the crawlers terms of the same books in the catalog of the Central Library of the University of Brasilia (CL/UNB). Presented as search results that text mining contributed significantly and reliable for pattern discovery for the thematic representation of information from the terms that appeared frequently and had semantic concordance in the text, making them candidates for further words key that will define the issues of the works. It is also, on the other hand, the text mining has not identified sufficient standards in two works, making it impossible perform the confrontation between the two environments. It is concluded that the comments made in web 2.0 have knowledge of users who enjoyed the work and made explicit his wisdom on posts, here transformed into keywords, showing that they can effectively participate in the thematic representation of information and further contribute to the success the search processes, allowing us to infer that for a possible faithful representation of the content of the work and better recovery, social indexing can support / complement traditional indexing, ie the tags of users, analyzed by mining, can be compared to the descriptors defined by librarians for making decision on the terms of representative works.

Keywords: Knowledge. Sapience. Thematic representation of information. Folksonomy. Data mining. Text mining.

LISTA DE ILUSTRAÇÕES

Figura 1	Tríade tradicional da sociedade	24
Figura 2	Tríade pragmática da sociedade	24
Figura 3	Ciclo informacional	26
Figura 4	Sabedoria na evolução do ciclo informacional	28
Figura 5	Elementos da tomada de decisão a partir da sabedoria	29
Figura 6	Ciclo da terceridade aplicada	30
Figura 7	Classificação dos seres vivos	36
Figura 8	Antigo diretório do <i>Yahoo</i>	37
Figura 9	Página colaborativa de usuário no <i>Delicious</i>	39
Figura 10	Página colaborativa de usuário no <i>Flickr</i>	39
Figura 11	Exemplo de rede social – <i>Facebook</i>	40
Figura 12	Representação de assuntos no <i>Twitter</i>	41
Figura 13	Exemplo de artigo na <i>Wikipédia</i>	41
Figura 14	Tela inicial de pesquisa na BRAPCI	44
Figura 15	<i>Website</i> do <i>Google Now</i>	47
Figura 16	Representação das etapas de KDD	49
Figura 17	<i>Software text minig suite</i>	53
Figura 18	Abordagens da mineração de dados na <i>web</i>	54
Figura 19	Tela do <i>software Websphinx</i>	55
Figura 20	Tela inicial do <i>software WEKA</i>	56
Figura 21	Biblioteca NLKT na função de classificação	57
Figura 22	Tela inicial do <i>software Kitconc</i>	61
Figura 23	Ferramenta de mineração – corpus	61
Figura 24	Ferramenta de mineração – lista de palavras	62
Figura 25	Ferramenta de mineração – lista de palavras	63
Figura 26	Ferramenta de mineração – concordância	63
Figura 27	Ferramenta de mineração – colocados	64
Figura 28	Ferramenta de mineração – N-gramas	64
Figura 29	Ferramenta de mineração – dispersão	65
Figura 30	Exemplo de livro no ambiente da <i>Amazon</i>	66
Figura 31	Catálogo da BCE/UnB	67
Figura 32	Obras escolhidas (OE) - parte 1	69
Figura 33	Obras escolhidas (OE) – parte 2	70
Figura 34	Comentários extraídos da obra “Os três porquinhos”	70
Figura 35	Corpus do livro	71
Figura 36	Coleta – lista de palavras	71
Figura 37	Coleta – palavras-chave	72
Figura 38	Coleta – concordância	73

LISTA DE QUADROS

Quadro 1	Ferramentas de mineração de dados	58
Quadro 2	Caracterização da pesquisa	59
Quadro 3	Análise do corpus de cada livro.	73
Quadro 4	Comparação de palavras-chave entre os ambientes	75

LISTA DE SIGLAS E ABREVIATURAS

AI	Arquitetura da Informação
BCE	Biblioteca Central
BRAPCI	Base de dados Referencial de Artigos de Periódicos em Ciência da Informação
CDD	Classificação Decimal de Dewey
CI	Ciência da Informação
DEMO	Versão de demonstração
DF	Distrito Federal
EUA	Estados Unidos da América
FURB	Fundação Universidade Regional de Blumenau
GPU	General Public License
IA	Inteligência Artificial
HTML	HyperText Markup Language
KDD	Knowledge Discovery in Database
MEC	Ministério da Educação
NLTK	Natural Language Toolkit
OE	Obras Escolhidas
PC	Personal Computer
SAB	Sistema de Automação de Bibliotecas
SO	Sistema Operacional
SRI	Sistema de Recuperação da Informação
TCC	Trabalho de Conclusão de Curso
TI	Tecnologia da Informação
TXT	Texto
UnB	Universidade de Brasília
UNISIST	United Nations International Scientific
URL	Uniform Resource Locator
XML	eXtensible Markup Language
WEKA	Waikato Environment for Knowledge Analysis

SUMÁRIO

1 INTRODUÇÃO	13
1.1 PROBLEMATIZAÇÃO	16
1.2 JUSTIFICATIVA	16
2 OBJETIVOS	18
2.1 OBJETIVO GERAL	18
2.2 OBJETIVOS ESPECÍFICOS	18
3 FUNDAMENTAÇÃO TEÓRICA	19
3.1 DOS DADOS À SABEDORIA: um ciclo de desenvolvimento e inovação	19
3.2 REPRESENTAÇÃO TEMÁTICA DA INFORMAÇÃO	31
3.2.1 Indexação: uma ação temática tradicional	33
3.2.2 Folksonomia: ação temática digital e colaborativa	36
3.3 MINERAÇÃO: ação de descoberta para decisão	46
4 METODOLOGIA	59
4.1 CARACTERIZAÇÃO DA PESQUISA	59
4.2 AMBIENTES DA PESQUISA E AMOSTRAGEM	65
5 ANALISANDO OS DADOS	69
6 CONSIDERAÇÕES FINAIS	77
REFERÊNCIAS	79

1 INTRODUÇÃO

As recentes mudanças tecnológicas fizeram com que a sociedade alterasse seu comportamento. Um indivíduo ao realizar, de forma cotidiana e instintiva, multitarefas no computador, acessando conteúdos diferentes e em ambientes diversos, insere-se em um contexto de consumo, produção e disseminação de um grande volume de informação, em um curto espaço de tempo. Este indivíduo, atualmente, encontra-se conectado digitalmente, fazendo uso não mais apenas dos tradicionais computadores, conhecidos como *personal computer* (PC)¹, mas também dos *tablets* e *smartphones*, promovendo ainda mais a supracitada produção de informação, independentemente da localização física. A citada produção é acessada e analisada pelas empresas e Instituições na perspectiva de dados, para a tomada de decisão, e assim proporcionar vantagens e lucros sobre seus produtos e serviços aos seus usuários.

Os dados são extraídos a partir de diversos recursos de armazenamento e gerenciamento de informações como, por exemplo, os serviços de *cloud computing*², de comunicação, de redes sociais etc. Como exemplo de um ambiente que integra os recursos anteriormente citados, podemos destacar o *Google*³ a partir do serviço de nuvem intitulado *Drive*⁴, de comunicação chamado *Hangout*⁵ e a rede social denominada *Google+*⁶ entre outros. Conforme as pesquisas dos usuários no seu buscador, o *Google* preenche o próprio banco de dados e conseqüentemente oferece resultados segundo o que foi digitado pelos mesmos, a fim de melhorar a qualidade das ocorrências, o *Google*, a partir da integralização dos seus serviços, toma decisões ainda mais precisas, levando em consideração a participação dos usuários em seus serviços quanto às características de perfil, necessidades e preferências. Como o *Google* usa os seus serviços para sugerir resultados de busca aos seus usuários, bem como os mesmos estão constantemente conectados, a grande massa informacional presente em seu banco de dados é analisada em tempo real. Desta forma o usuário receberá respostas também em tempo real, deflagrando-se uma análise de dados intitulada *big data*, ferramenta definida como a descrição de um conjunto de problemas e soluções tecnológicas com base em formas inovadoras de processamento de informação para a tomada de decisão em grandes volumes de dados.

¹ Computador Pessoal.

² Computação nas nuvens.

³ Website: <https://www.google.com.br/>.

⁴ Website: <https://drive.google.com/>.

⁵ Website: <http://www.google.com/+learnmore/hangouts/?hl=pt-br>.

⁶ Website: <https://plus.google.com/?hl=pt-BR>.

De acordo com Somasundaram e Shrivastava (2011, p. 21), “[...] enormes quantidades de informações digitais são criadas a todo momento por consumidores individuais e corporativos de TI⁷. Esses dados precisam ser armazenados, protegidos, otimizados e gerenciados”.

Muitas das informações que nos permeiam hoje nascem e/ou são apresentadas na *web*, obtidas em *websites*, *blogs*, redes sociais etc., constituindo, desta forma, uma nova forma de produção do conhecimento humano. Devido a esta disponibilização torna-se ainda mais importante a necessidade de organizar essa informação com fins de recuperação e uso. A referida ação, de ordenar a informação, foi sendo moldada conforme a evolução da *web*. Nos primórdios, as informações da *web* eram estruturadas hierarquicamente, onde o produtor e/ou administrador do ambiente tinham o total controle na forma como eram produzidas, editadas, organizadas e disseminadas. Esta primeira geração, intitulada de *web* 1.0, tinha o caráter comercial, pois os usuários somente usufruíam as informações na posição de leitor conforme predefinições realizadas pelos proprietários dos *websites*, entretanto vale salientar que ainda existem muitos ambientes da *web* com este formato. Ainda nesta primeira fase, os usuários navegavam por *links*, comumente organizados em diretórios de assuntos, para encontrar a informação desejada. Por outro lado, como a quantidade de informação disponível não era semelhante ao panorama atual e a tipologia era restrita, geralmente os resultados de busca eram mais precisos. Em suma, a navegação era vista como um ato exaustivo, enquanto os resultados de busca eram considerados como específicos.

Com o tempo surgiu uma *web* que permitiu ao usuário participar da representação e organização da informação no respectivo ambiente. A partir desta participação, cada vez mais ativa e frequente dos usuários, o volume de informações geradas na *web* cresceu consideravelmente, apresentando a *web* 2.0, fase “[...] colaborativa por natureza, interativa, dinâmica, e a linha entre criação e consumo de conteúdo nesses ambientes era tênue (usuários criavam o conteúdo nesses *websites* tanto quanto eles o consumiam)” (MANESS, 2007, p. 43). Entre suas características, a marcante chama-se folksonomia, ou seja, uma maneira pela qual os usuários podem representar as informações de maneira livre, por meio das *tags*, e posteriormente classificá-las sem controles terminológicos e de arranjos. Conforme Catarino (2007, p. 3), “folksonomia é o resultado da atribuição livre e pessoal de etiquetas (*tagging*) a informações ou objetos (qualquer coisa com *URL*⁸), visando à sua recuperação. A atribuição de etiquetas é feita num ambiente social (compartilhado e aberto a outros)”.

⁷ Tecnologia da Informação.

⁸ *Uniform Resource Locator*.

A folksonomia retrata a sabedoria da sociedade, diferentemente do que acontece com a indexação enquanto um ato de representação temática da informação na perspectiva das bibliotecas físicas. Enquanto a primeira ocorre a partir dos termos atribuídos pelas pessoas na *web*, a segunda é realizada segundo a subjetividade dos profissionais da informação, como os bibliotecários. Enquanto a folksonomia retrata o saber popular, a indexação volta-se ao conhecimento do profissional, ampliando o debate sobre a tradicional tríade da sociedade da informação (dado, informação e conhecimento) conforme a inclusão da sabedoria neste cenário discursivo.

Ao verificar um grande volume de usuários nos ambientes atuais da *web* torna-se recomendável identificar possíveis padrões quanto às formas que estes produzem, consomem, disseminam e compartilham informações com fins de apresentar melhores sugestões em suas buscas, como também auxiliar na definição de termos representativos (indexadores) aos mais variados produtos. Chega-se ao discurso, com base na compreensão de Nicholson (2004, p. 254), de que “[...] é na descoberta desses padrões que se encontra a chave para conhecer melhor as comunidades de usuários dos serviços oferecidos pelas bibliotecas, com o objetivo de melhor adequar estes serviços aos diferentes grupos de usuários”. Nesse sentido, o presente trabalho caminha no vislumbre de conhecer a mineração de dados e suas potencialidades de aplicação no campo da biblioteconomia. Ainda de acordo com Nicholson (2004, p. 260), esta ferramenta trata “[...] da descoberta de padrões de dados em nível inferior, utilizando ferramentas de estatísticas e de inteligência artificial”.

A mineração de dados afigura-se como uma técnica de análise que pode ser utilizada no universo biblioteconômico, mais especificamente na representação temática da informação, confrontando termos atribuídos na folksonomia com palavras-chave definidas na indexação tradicional. Assim a biblioteconomia pode identificar novas oportunidades de atuação, especialmente nos ambientes digitais, e aproveitá-las a fim de conquistar mais usuários, conhecendo-os e tratando-os de forma singular, ou seja, segmentando-os conforme suas preferências e indicando os materiais que mais se adequem às necessidades informacionais segundo os termos atribuídos pelos próprios ao representar algum objeto, entre livros e multimeios. De acordo com Medeiros, Da Silva e Brito (2014, p. 135), os multimeios “são suportes de informação constituídos de materiais não convencionais, que possuem, em sua característica, a marca do registro histórico e da evolução constante, uma vez que, a cada dia, surge uma nova tecnologia”.

A mineração de dados como uma possibilidade de ferramenta no subsídio à tomada de decisão em ações como representação temática da informação e estudos de usuários. Enquanto a primeira contribuirá na escolha de palavras-chave e definição dos assuntos das obras, a segunda ação ajudará na busca da identificação do perfil e interesses dos usuários.

Diante do contexto apresentado este estudo está estruturado da seguinte maneira: o capítulo inicial apresenta a introdução, a problematização e a justificativa; no capítulo dois são elencados os objetivos da pesquisa; no capítulo três está dissertada a fundamentação teórica, abordando elementos que compõem o ciclo de desenvolvimento e inovação da sociedade, a representação temática da informação, dando ênfase na indexação tradicional e na folksonomia, e a mineração de dados, com destaque a análise de textos; no capítulo quatro delimita-se a metodologia do presente estudo; no capítulo cinco são analisados os termos indexadores oriundos de textos provenientes da folksonomia, segundo a mineração de textos, e posterior comparação de termos entre a indexação social e a tradicional, extraídos de um ambiente digital e outro físico respectivamente; e o último capítulo reflete-se as considerações finais e aspirações para estudos futuros.

1.1 PROBLEMATIZAÇÃO

O presente trabalho caminha na proposição de uso da mineração de dados enquanto elemento contribuidor na tomada de decisão, mais especificamente no processo de representação temática da informação ao analisar a participação do usuário a partir da folksonomia em ambientes da *web 2.0*. Dado este contexto, a pergunta que instiga este trabalho de conclusão de curso (TCC) está assim formulada: *de que forma a mineração de dados pode contribuir na representação temática da informação?*

1.2 JUSTIFICATIVA

Com participação ativa e frequente dos usuários nos recentes ambientes da *web*, o volume de informações geradas cresceu consideravelmente. Neste sentido, enquanto usuária e produtora de informações na *web*, vislumbramos a folksonomia como um aperfeiçoamento da representação temática da informação, precisamente na indexação tradicional, tendo a sabedoria dos usuários como ponto de partida para a descoberta de padrões a partir das análises realizadas com as técnicas de mineração de dados. A citada visão nos permitiu construir, ao longo das disciplinas cursadas durante a graduação em biblioteconomia, o problema de

pesquisa. Durante este período, ao participar de uma disciplina chamada “Seminário em Biblioteconomia”, no semestre 1/2012, do curso de Biblioteconomia, da Universidade de Brasília (UnB), conhecemos a mineração de dados, assunto de uma determinada palestra da disciplina. Guardei a informação adquirida, percebendo que poderia ser algo a ser estudado e talvez uma boa ferramenta a ser usada em nossa área. Passados alguns semestres, especificamente enquanto aluna da disciplina obrigatória intitulada “Redes de Informação e Transferência de Dados”, no 1/2013 semestre, foi possível perceber aproximações entre o tema *web* 2.0, ao qual se falava da identificação de padrões na folksonomia, com a mineração de dados, especialmente pelas possíveis contribuições deste tipo de análise na representação temática da informação enquanto uma forma de trabalho não habitual ao campo da biblioteconomia.

2 OBJETIVOS

2.1 OBJETIVO GERAL

- Analisar a mineração de dados como ferramenta para a tomada de decisão na representação temática da informação.

2.2 OBJETIVOS ESPECÍFICOS

- Elencar os elementos que formalizam um ciclo de desenvolvimento e inovação da sociedade;
- Apresentar a representação temática da informação nas vertentes física e digital;
- Verificar os tipos de mineração de dados;
- Analisar o uso da mineração de textos na folksonomia com fins de identificação de padrões;
- Confrontar o resultado da mineração de textos na folksonomia com termos da indexação tradicional.

3 FUNDAMENTAÇÃO TEÓRICA

Nesta seção é apresentada uma parte fundamental da pesquisa científica, que é o levantamento de teorias e conceitos para embasar todos os objetivos propostos pelo pesquisador, isto é, a construção da base da pesquisa, um quadro teórico para a estruturação conceitual e desenvolvimento da pesquisa. Também chamada como o “estado da arte”, a fundamentação teórica é o estrato essencial para se estabelecer a conjuntura dos temas pesquisados no problema de pesquisa, alcançar questões ainda não respondidas, além de adquirir e/ou aprofundar o conhecimento sobre o assunto abordado. Bento (2002, p. 42) cita a fundamentação teórica enquanto uma revisão de literatura, a ação “[...] indispensável não somente para definir bem o problema, mas também para obter uma ideia precisa sobre o estado atual dos conhecimentos sobre um dado tema”.

Em suma, a presente fundamentação teórica parte de discussões conceituais formalizadas pelos temas dado, informação, conhecimento e sabedoria; representação da informação, em sua perspectiva temática; e a mineração de dados, como uma ferramenta de análise que pode contribuir no processo de representação temática.

3.1 DOS DADOS À SABEDORIA: um ciclo de desenvolvimento e inovação

No momento o homem percebeu que não seria capaz de calcular a quantidade dos seus animais e objetos pessoais somente com o auxílio dos dedos das mãos, gravetos e/ou pequenas pedras, surgiu a primeira ideia do conceito que chamamos de dado. É a partir dessa ação que “máquinas” de calcular foram construídas ao longo de 4.000 a.C., iniciadas pelo ábaco, que por sua vez foi a base para a pascalina chegando a máquina de contar com os cartões perfurados. Tempos depois, Charles Babbage considerado o pai da computação, construiu um instrumento que faria alguns cálculos avançados utilizando os cartões perfurados, chamada máquina diferencial. A partir da inovação da Pascalina e de Babbage, qualquer coisa que possa ser estruturado e capaz de ser lido por um computador é considerado um dado, definido de acordo com Setzer (1999, p. 2, grifos do autor) da seguinte forma:

Dado é necessariamente uma entidade matemática e, desta forma, puramente *sintática*. Isto significa que os dados podem ser totalmente descritos através de representações formais, estruturais. Dentro de um computador, trechos de um texto podem ser ligados virtualmente a outros trechos, por meio de contiguidade física ou por "ponteiros", isto é, endereços da unidade de armazenamento sendo utilizada.

Dados estão particularmente ligados ao computador, corroborando à ideia de que é parte fundamental para a construção de uma informação. O autor mencionado ainda afirma que dados são uma sequência de símbolos quantificados e quantificáveis. Ao utilizar nossos cinco sentidos, sem agregar valor à ação que estamos realizando, seja ouvindo, cheirando, observando, tocando ou saboreando, estamos obtendo dados relacionados a essas ações no nosso cérebro; Davenport e Pruzak (1999, p. 18) observam que “dados são simples observações sobre o estado do mundo, são facilmente estruturados, obtidos por máquinas, frequentemente quantificados e facilmente transferidos”. Os dados têm um importante papel enquanto agente para a compreensão de fenômenos da natureza e do conhecimento científico. Corroboram com esta ideia Santos e Sant’Ana (2002, p. 5) ao afirmarem que:

[...] os dados, permitem apenas inferências baseadas em informações complementares que poderiam estar relacionadas àquele dado. Porém, o dado em si, se transmitido ou registrado fora do seu contexto, pouco ou nada pode representar em termos de significado.

É possível compreender que os dados sem contexto são apenas elementos sem sentido, itens desestruturados em um sistema, por vezes não tratados em um ambiente bibliográfico. Neste sentido, na indagação sobre o que pode ser considerado, ou não, dado na vida do homem, Oliveira (2005) defende que:

[...] dados têm a potencialidade de carregar informação, mas é necessário partilhar de uma forma de vida para ser capaz de distinguir o que é ou não um dado. Isto pode ser algumas vezes difícil de perceber, pois vivemos em um contexto social onde partilhamos uma forma de vida que nos capacita distinguir qual dado contém essa possibilidade de informação e qual não, e, ao mesmo tempo nos torna cegos para o fato de que nós, somente podemos distinguir entre esses dados, porque partilhamos uma cultura, ou seja, uma forma de vida.

Quando existe agregação de valores, como o ambiente onde estão inseridos, ou seja, um fato histórico, por exemplo, os dados tornam-se passíveis de interpretação/compreensão, configurando-se em um processo de criação de informação. Esta valoração transforma os dados em informação para os receptores, variando a compreensão conforme o conhecimento de cada um. Nesse sentido Furgueri (2006, p. 27) afirma que “a informação se refere à interação dos dados para produzir ativação e significado”. Francelin e Pellegatti (2004) corroboram com a discussão ao defenderem que os dados são essenciais para a construção da informação, assim sendo o dado um molde inicial na construção de algo.

O presente debate, sobre a importância do dado e consequente produção de informação, gera interesses de pesquisas, tornando a própria informação objeto de estudo de várias áreas do conhecimento humano, justificado pelo fato de ser um fator importante para a tomada de decisão, representar à realidade humana e, essencialmente, um motor de desenvolvimento das sociedades.

Entre tais práticas de estudo podemos citar a proposta do estudioso Shannon Weaver, que desenvolveu um modelo de como resultava a troca de informação, chamada de Teoria da Informação, também conhecida por Teoria da Comunicação, com o objetivo de descrever como ocorria a transferência da informação no modelo emissor-canal-receptor. Inserido no contexto da II Guerra Mundial, o que mais importava para Shannon eram os termos técnicos dessa transferência, ou seja, como a sua teoria poderia melhorar a emissão da mensagem e não a semântica. Com isso, na área da comunicação, conforme o avanço desta teoria aprimorou-se o modelo para investigar os meios pelos quais a informação transita e de que forma é utilizada na sociedade, pesquisando os efeitos positivos que a mesma sem ruídos possuía. No caso da linguística, busca-se descobrir como a informação pode influenciar na construção de linguagens a partir de estudos advindos da semiótica, que é uma ciência que estuda os signos e como os mesmos, e significados, podem influenciar uma cultura ou comunidade. Na filosofia almeja-se conceituá-la e investigar seus princípios básicos enquanto promotores do conhecimento humano, não se preocupando com o seu sentido pragmático.

Para Saracevic (1974, p. 64) “[...] o fenômeno da informação é estudado em muitas disciplinas diversas confirmando assim as ramificações complexas e muitas manifestações a ele associadas”. Nesse sentido, cada área limita-se a elaborar um conceito de informação. Apesar dos enfoques em cada área supracitada, Le Coadic (1996, p. 5) une vários conceitos:

[...] conhecimento inscrito (gravado) sob a forma escrita (impressa ou numérica), oral ou audiovisual. A informação comporta um elemento de sentido. É um significado transmitido a um ser consciente por meio de uma mensagem inscrita em um suporte espacial temporal: impresso, sinal elétrico, onda sonora, etc. Essa inscrição é feita graças a um sistema de signos (a linguagem), signos estes que são elementos da linguagem que associa um significante a um significado: signo alfabético, palavra, sinal de pontuação.

Assim como os dados necessitam de contexto e valores para se transformarem em informação, esta última, ao longo do tempo, foi conceituada de acordo com as considerações de cada área, conforme pôde ser observado nas áreas da comunicação, linguística e filosofia anteriormente elencadas. Mesmo com as peculiaridades de cada área, todas levam em

consideração o processo de transferência da informação, onde emissor e o receptor são essenciais para a modificação do dado em informação, que por sua vez se transformam em conhecimento.

[...] A informação, na verdade, é indispensável para toda e qualquer atividade humana, sendo, cada vez mais, vista como uma força importante e poderosa a ponto de dar origem a expressões como: sociedade da informação, explosão da informação, era da informação, indústria da informação, revolução da informação, sociedade pós-sociedade da informação. (ARAÚJO; FREIRE, 1999, p. 5)

A tríade (dado, informação e conhecimento), formalizada nas Eras supracitadas e tradicionalmente discutida na literatura, apresenta rastros de discussão na biblioteconomia, campo de atuação da Ciência da Informação (CI) que lida com a informação de maneira enfática, seja ela gravada, inscrita e explícita em algum suporte. Para tanto citamos o pai da biblioteconomia moderna, o indiano Shiyali Ramamrita Ranganathan (1892-1972), teórico que formulou as cinco leis da biblioteconomia: a primeira, os livros são para serem lidos; a segunda, para cada leitor seu livro; a terceira, cada livro o seu leitor; a quarta, poupe o tempo do leitor; e a quinta, a biblioteca é um organismo em crescimento. Interpretando essas leis, na discussão sobre o enlace entre dado e informação, travado até aqui, no momento em que cada item de uma obra é tratado descritivamente no sistema de uma biblioteca, este representa um dado. Em outro momento, ou seja, quando, o leitor tem a percepção de que o conteúdo daquele item preencherá as suas lacunas intelectuais, este item passa a se transformar em informação. Neste instante, em que a biblioteca verifica que tal item foi acessado, temos uma informação que pode tornar-se um elemento-chave para a tomada de decisão. Ao utilizar a informação, o leitor estará produzindo e/ou ampliando o seu conhecimento, embasando-o na decisão quanto ao uso de outras fontes e serviços da biblioteca. Neste sentido, a biblioteca toma como posse as ações dos usuários para decidir sobre a ampliação do acervo e melhorias dos serviços prestados, justificando assim todas as leis de Ranganathan, em especial a última, que apresenta a biblioteca como um espaço que vive em constante de crescimento não possuindo a visão das cinco leis, as coleções serão encaradas apenas como dados, tanto pelo profissional que as gerencia, como também pelos usuários. Quando as coleções são organizadas conforme os objetivos propostos pela Instituição e seguindo regras de arranjo, o acervo estará apto aos usuários, ou seja, estará ordenado pelo bibliotecário conforme algum sentido. Já na visão do usuário, ao usar um acervo devidamente organizado, os livros deixarão de ser apenas dados, passando para suportes que

apresentam informações indispensáveis às suas necessidades, que por sua vez afetarão diretamente no próprio conhecimento.

Podemos observar que a informação, no âmbito da biblioteconomia, está geralmente registrada em algum suporte e que pode ser útil a algum indivíduo, encarado como um usuário (leitor) em potencial, capaz de ler àquele conjunto de dados. Wurman (1995, p. 48) compreende que informação existe quando pode ser entendida por alguém ou algo, é "aquilo que leva à compreensão que constitui informação para uma pessoa pode não passar de dados para outra". Conforme Cedón *et al.* (2005, p. 18), informação “[...] é algo que ajuda na resolução de um problema ou completa uma lacuna no conhecimento da pessoa, conforme cada necessidade”. Nesta situação, os produtores de informação procuram preencher suas próprias necessidades de conhecimento, bem como podendo suprir as lacunas de outros ao publicar/disseminar informação, seja transmitida de forma oral, manual ou digital.

Assim como ocorre na relação entre dado e informação, esta última também passa por um processo de transformação, gerando o conhecimento. Essencialmente humano, tanto social como individual, o conhecimento é desenvolvido pelas informações captadas do próprio ser durante sua evolução e conseqüente adaptação às características da sociedade presente.

[...] a diferença entre informação e conhecimento seja mais como uma diferença de graus em uma escala do que uma distinção de categoria. Informação e conhecimento são entendidos como produtos da inteligência humana e do linguajar, e sua capacidade de selecionar, nomear, ordenar e juntar. Informação e conhecimento são parte de um mesmo processo. Processo este que inclui a capacidade de arrumar e rearrumar padrões, usando elementos e grupos de elementos; como um processo contínuo de transformação, no qual as transformações são desencadeadas por agentes perturbadores (internos ou externos) e determinadas por nossa estrutura e sua contínua história de transformações [...]. (OLIVEIRA, 2005, p. 5)

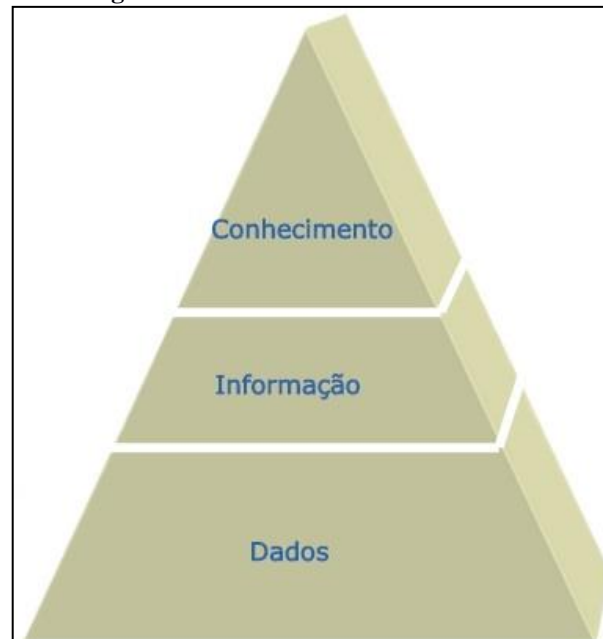
Para alguns autores, como Barreto (2002, p. 68, grifos do autor), o conhecimento dificilmente será conceituado por se tratar de um fluxo muito subjetivo do ser humano:

Entende-se o conhecimento como um fluxo de acontecimentos, isto é, uma sucessão de eventos, que se realizam fora do estoque, na mente de algum ser pensante e em determinado espaço social. É um caminho subjetivo e diferenciado para cada indivíduo. Daí ser incompreensível os conceitos de “gestão do conhecimento” ou “base de dados de conhecimento”.

De forma ampla, o conhecimento pode ser apresentado como um produto, um processo, uma transformação ou ainda como um fluxo de evolução, inversamente proporcional, dos dados e da informação, conforme apresentado na figura um (1). A referida ilustração pode ser

considerada como uma ilustração que representa a tríade tradicional que constitui a sociedade da informação, onde os dados são os elementos basilares para a construção de informações e assim, conseqüentemente, o conhecimento é o produto de várias informações. De forma ampla, a partir desta dinâmica, o conhecimento torna-se um produto essencial para o processo decisório nas organizações.

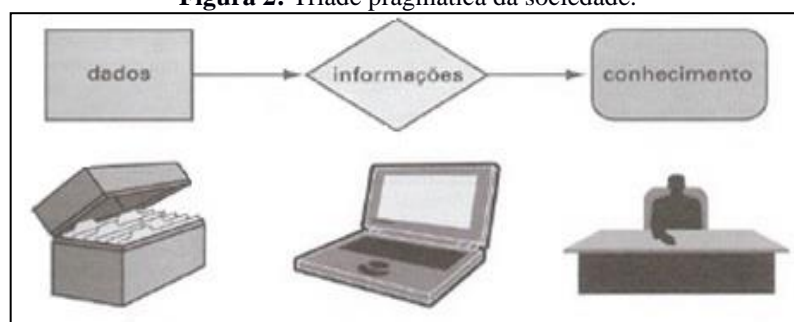
Figura 1: Tríade tradicional da sociedade.



Fonte: Canudo, 2007.

Complementando a discussão, os dados são fatos básicos que podem ser estruturados, quantificados, transferíveis e lidos por máquinas. Quando inseridos em um determinado contexto, pela atividade humana dotada de significados, podem ser transformados em informações, que por sua vez geram conhecimentos, entre tácito e explícito. A citada dinâmica é mencionada por Gordon e Gordon (2006) como uma tríade pragmática, ilustrada na figura dois (2).

Figura 2: Tríade pragmática da sociedade.



Fonte: Gordon e Gordon (2006) apud Canudo, 2007.

O conhecimento humano é gerado a partir de um conjunto de informações, que foram sendo adquiridas em um determinado período de vida. Entretanto, para que esse processo aconteça em sua plenitude, Oliveira (2005, p. 8) defende que o conhecimento precisa ser aplicado para a sua efetiva formalização ao afirmar que:

Conhecimento implicaria em aplicação, ou possibilidade de aplicação, a informação seria um conhecimento sem aplicação. É por isso que informação passa a [ideia] de ser capaz de ser armazenada, enquanto é mais difícil pensar em conhecimento sendo armazenado. Para que seja possível desencadear, a partir de um dado codificado, uma informação em alguém, o que é requerido é que se partilhe de uma mesma forma de vida.

Conforme a aplicação, encontramos dois tipos de conhecimento: o tácito e o explícito. De forma objetiva, o tácito é aquele que está na mente humana, enquanto o explícito é o conhecimento que registramos em qualquer tipo de suporte.

[O conhecimento tácito] é aquele que não tem como ser totalmente convertido em um conjunto de códigos ou sinais, não permitindo sua transmissão completa e, portanto, não podendo ser registrado em meios artificiais, tendo sua existência, ligada diretamente às pessoas que o detêm. (SANTOS; SANT'ANA, 2002, p. 6)

Portanto o tácito é aquele conhecimento em que não temos a capacidade de registrar, ao qual apresentamos somente com ações, diferentemente do explícito, definido como:

O explícito corresponde à parte do conhecimento que pode ser convertido em um conjunto de informações e, portanto, ser transmitido e conseqüentemente registrado em suportes artificiais, e representa apenas uma pequena parte de todo o conhecimento, aquela que pode ser formalmente articulada de maneira mais precisa. Ele pode ser mais facilmente codificado, documentado, transferido ou compartilhado. (SANTOS; SANT'ANNA, 2002, p. 6)

Após os conceitos elencados sobre a tríade discutida, Setzer (1999, p. 4) apresenta-nos a relação entre esses elementos, ou seja, entre dado, informação e conhecimento:

[...] o conhecimento não pode ser descrito inteiramente - de outro modo seria apenas dado (se descrito formalmente e não tivesse significado) ou informação (se descrito informalmente e tivesse significado). Também não depende apenas de uma interpretação pessoal, como a informação, pois requer uma vivência do objeto do conhecimento. Assim, quando falamos sobre conhecimento, estamos no âmbito puramente subjetivo do homem ou do animal.

Levando a discussão para o contexto do armazenamento, Setzer (1999, p. 6) apresenta o seguinte extrato sobre os elementos dado, informação e conhecimento:

Os dados que representam uma informação podem ser armazenados em um computador, mas a informação não pode ser processada quanto a seu significado, pois depende de quem a recebe. O conhecimento, contudo, não pode nem ser inserido em um computador por meio de uma representação, pois senão foi reduzido a uma informação.

Complementando a presente tentativa de estratificar a tríade, Fernandez-Molina (1994, p. 328, tradução nossa) a resume conceituando da seguinte forma:

Dados, informação e conhecimento: dados são informação em potencial, que somente são percebidos por um receptor se forem convertidos em informação e esta passa a converter-se em conhecimento no momento em que produz uma modificação na estrutura do conhecimento do receptor.

A tríade também pode ser discutida enquanto um ciclo informacional, onde a informação é a base para o conhecimento, proporcionando insumos para o desenvolvimento de determinada área em um movimento contínuo (figura 3).

Figura 3: Ciclo informacional.



Fonte: Barreto (1998, p. 123).

Barreto (1998) apresenta outra abordagem referente a tríade tradicional da sociedade, aqui intitulada de ciclo informacional, ao qual gera desenvolvimento, em um processo contínuo, a partir de informações e conhecimentos, formalizado por Oliveira (2002, p. 38) da seguinte maneira:

O desenvolvimento deve ser encarado como um processo complexo de mudanças e transformações de ordem econômica, política e, principalmente, humana e social. Desenvolvimento nada mais é que o crescimento [...] transformado para satisfazer as mais diversificadas necessidades do ser humano, tais como: saúde, educação, habitação, transporte, alimentação, lazer, dentre outras. [...] É desta maneira que o desenvolvimento passa a ser entendido como uma resultante do processo de crescimento, cuja maturidade se dá ao atingir o crescimento [auto-sustentado], ou seja, talvez alcançar a capacidade de crescer sem fim, de maneira contínua.

O desenvolvimento é parte importante do ciclo, pois se refere ao crescimento, ao proporcionar novas informações. Este é “simultaneamente contínuo, (ocorre gradual e quantitativamente de forma cumulativa e aditiva) e descontínuo condicionando novos comportamentos e formas de pensamento” (VALE, 2005, p. 105).

A tríade oferecida por Barreto (1998) pode ser considerada como uma geométrica basilar às inovações, ou seja, algo inédito gerado a partir do conhecimento obtido, através das informações processadas anteriormente. De acordo com Setzer (1999), a inovação vem da habilidade de aplicar o conhecimento humano em algo no “mundo real”, que necessita do conhecimento e capacidade pessoal. Nicolau e Paranhos (2006, p. 34) vão além afirmando que “a inovação é o resultado coletivo de muitos atores, interagindo e formando um sistema”.

Outro elemento que vem ampliando a tradicional tríade da sociedade é a sabedoria, devido ao recente momento de produção coletiva e compartilhamento de informação nos ambientes digitais, especialmente na *web*. Silva e Santos (2008) compreendem que a sabedoria está ligada ao aprendizado organizacional, à geração de novos conhecimentos e o desenvolvimento de uma cultura organizacional voltada à experimentação e aprendizado contínuo, comprometida com os resultados que a organização atribui aos seus colaboradores.

A sabedoria desenvolve-se a partir de interações e colaborações entre indivíduos, seja no âmbito profissional, como também na comunicação científica e atividades de lazer. A interação e a colaboração com a troca de conhecimentos, gerando o que chamamos de inteligência coletiva, são importantes para o processo de produção de inovações das Instituições. Deste modo, conforme os atuais ambientes digitais, o ciclo informacional, proposto por Barreto, evolui para o adicionamento da passagem do conhecimento para a inteligência coletiva, ou seja, a sabedoria.

Figura 4: Sabedoria na evolução do ciclo informacional.

Fonte: Google imagens, 2014.

A figura quatro (4) representa uma forma de ilustrar a evolução da tríade tradicional, com o acréscimo da inteligência coletiva advinda da ação do conhecimento explícito. Neste contexto, Marteleto (2007) nos insere no mundo da globalização, a partir de uma visão de terceiridade da informação, apresentando-nos o conceito de sabedoria adquirida a partir de grupos, ou seja, redes sociais físicas.

Ao se relacionarem os conceitos de informação, rede, conhecimento e saber pode-se imaginar que o saber, na sua organização abstrata e geral, toma a forma de uma rede, à qual as tecnologias modernas de organização e arquivamento podem conferir uma realidade concreta e palpável. No entanto o saber, enquanto conjunto de conhecimentos teóricos reunidos em diferentes *corpus* não representa mais do que uma ínfima parte do sistema geral de produção de conhecimentos. (MARTELETO, 2007, p. 8)

Seguindo a lógica da figura quatro (4), a sabedoria tem a dimensão do agir, um campo que permite a reflexão baseado na experiência de vida que os indivíduos possuem nestas redes sociais. Sendo assim, a terceiridade da informação ocorre a partir do conhecimento e da sabedoria interligados mutuamente, resultando em ações de interação, interatividade e sabedoria. Interativo requer a coletividade de conhecimentos e a iteratividade necessita de reflexão no momento antes de agir, podendo voltar em qualquer etapa anterior, ou seja, acrescentar outras informações ou conhecimentos. A sabedoria é a ação reflexiva dos conhecimentos adquiridos durante o tempo da vida, explícito ou não, que espelha a plenitude do indivíduo inserido nas redes sociais.

Num encadeamento didático dos conceitos de informação-conhecimento-saber, a informação estaria situada num primeiro grau do processo de apropriação, expressão e sistematização dos significados, enquanto o conhecimento corresponde a um grau mais internalizado desse mesmo processo, o de produção de sentidos sobre as coisas e o mundo. O saber situa-se numa terceira zona que reúne os cabedais sociais, técnicos e cognitivos de instituições, pessoas e grupos, em grau de internalização e externalização que alia os conhecimentos às práticas e às experiências, formando acervos, registros e documentos individuais e coletivos – a sabedoria. (MARTELETO, 2007, p. 10)

Tendo como produto da sabedoria, a tomada de decisão passou a receber elementos constituídos pelas pessoas da sociedade, pelos usuários dos ambientes digitais, como é o caso da representação e organização da informação na *web 2.0*, a partir de uma cadeia que ocorre conforme a figura (5). Nesta fase da *web*, os ambientes são usados e atualizados conforme a participação do usuário, em ações de navegação, acesso, criação, edição, disseminação e compartilhamento de informação, ocorrências que ocorrem nos serviços integralizados do *Google*, mencionados na introdução deste trabalho (seção 1).

Figura 5: Elementos da tomada de decisão a partir da sabedoria.



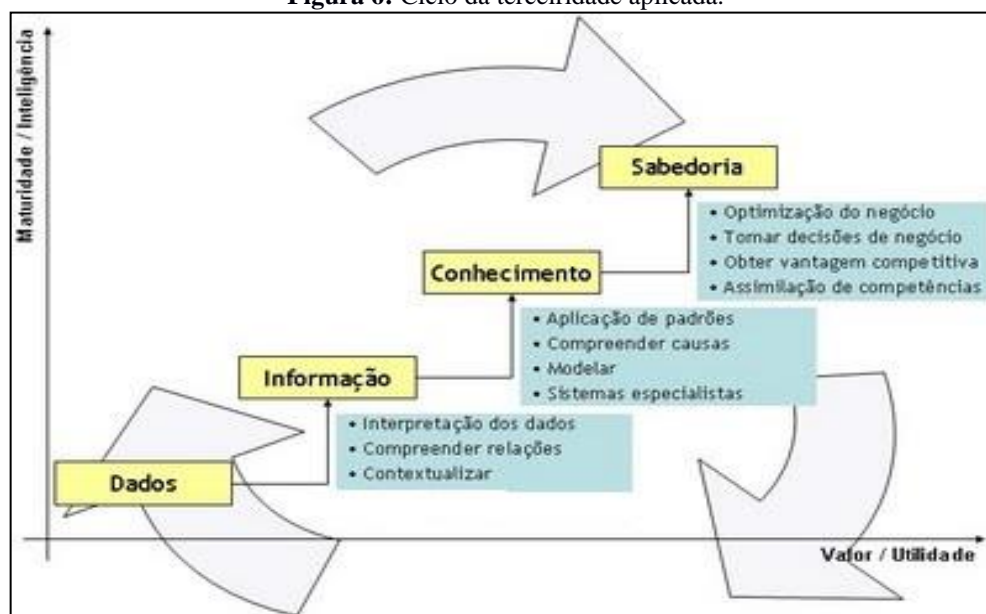
Fonte: Google imagens, 2014.

Em suma, a informação e o conhecimento bem utilizados são fundamentais para o processo de inovação, ativando as capacidades de desenvolvimento e sabedoria no ambiente em que se encontra o indivíduo, assim como afirmam as autoras:

A informação e o conhecimento estão em todas as esferas e áreas, são considerados essenciais tanto do ponto de vista acadêmico quanto profissional e, quando transformados pelas ações dos indivíduos, tornam-se competências valorizadas, gerando benefícios sociais e econômicos que estimulam o desenvolvimento. (TOMAEL; ALCARÁ; DI CHIARA, 2005, p. 93)

De forma geral, conforme os conceitos apresentados, é possível compreender que os dados sem contexto, sem cultura, podem ser perdidos em um sistema e assim não contribuem na produção de informação. Já se forem inseridos em algum espaço, um momento, uma situação, os dados podem ser contextualizados e, conseqüentemente, subsidiarem a produção de informação. No caso da informação é requerido um emissor-canal-receptor para completar seu objetivo de transferência de mensagens entre os envolvidos. Logo que se completa esta etapa, de transferência, o indivíduo mentalmente produz conhecimento. Este produto da informação pode ficar somente na perspectiva tácita ou ser explicitado em algum suporte. Quando o conhecimento explícito é posto em prática, é gerado um desenvolvimento institucional e social. Os dois âmbitos, nos tempos atuais, vêm recebendo influências da coletividade social, promovendo a ampliação do tradicional discurso triangular (dado, informação e conhecimento) com a inserção da sabedoria nos processos de tomada de decisão, delineando o ciclo da terceiridade aplicada (figura 6).

Figura 6: Ciclo da terceiridade aplicada.



Fonte: Google imagens, 2014.

Portanto a sabedoria permitirá que o ciclo recomece a partir de novas produções de conhecimentos promovidos da interação entre grupos sociais, constantes nos ambientes digitais da *web 2.0*. É neste contexto que a representação temática da informação acontece, ação que coloca uma coisa no lugar de outra conforme significados que fazem sentido as comunidades assistidas, inicialmente estudadas nos ambientes físicos e cada vez mais examinadas e aplicadas nos espaços digitais.

3.2 REPRESENTAÇÃO TEMÁTICA DA INFORMAÇÃO

A representação temática da informação faz-se necessária para o contínuo processo do ciclo informacional, uma vez que ao representar a informação, ela propiciará a recuperação da mesma, fomentando e atendendo, assim, a necessidade de conhecimento. Na CI, especialmente na biblioteconomia, ações são realizadas para que as informações sejam disponibilizadas aos usuários a partir de pontos de acesso, *links*, remissivas etc. compreensíveis à linguagem dos mesmos. A este conjunto de atividades intitulamos de tratamento da informação, ao qual, de acordo com Dias (2001) “[...] é definido como a função de descrever os documentos, tanto do ponto de vista físico (características físicas dos documentos) quanto do ponto de vista temático (ou de descrição do conteúdo)”. Até os anos 80, o tratamento da informação era tido como ação única da catalogação, atividade que abarca toda a descrição do documento.

[Catalogação] é o processo por meio do qual se descreve formalmente um documento ou recurso e se estabelece um variado e variável número de pontos de acesso, objetivando proporcionar, ao usuário final, a condição de encontrar, identificar, selecionar e obter o documento ou o recurso descrito, ou a informação nele contida. (CATARINO; SOUZA, 2012, p. 84)

Tempos depois, a compreensão sobre o tratamento da informação foi sendo reconstruído, assim como defende Mey (1987) ao afirmar que processos como a indexação e a classificação, que fazem parte da catalogação, desenvolveram-se largamente com os avanços da teoria da informação. A partir desse momento, de acordo com a literatura revisada, a catalogação descritiva se tornou representação descritiva e a catalogação de assunto que envolve a indexação e a classificação passam a ser representação temática.

Diante da atualização passou-se a usar a denominação chamada representação da informação. Para Guedes e Dias (2010, p. 41) a “[...] é um meio de expressão, uma linguagem na qual se pode dizer coisas sobre o mundo”, ou seja, algo que possa transmitir um conceito ou determinado assunto a uma comunidade, respeitando suas características sociais, culturais e linguísticas. Trata-se de “[...] um processo que pode utilizar linguagem específica uma vez que possui características próprias e seu principal objetivo é proporcionar a comunicação entre os documentos e os usuários da informação” (MAIMONE; SILVEIRA; TÁLAMO, 2011, p. 28). Verifica-se que os usuários são parte importante no processo, pois estes devem entender a mensagem que representa determinado documento, certa informação. A representação da informação é um processo cognitivo básico que surge desde o momento em que classificamos alguma coisa em nossa mente, gerando conceitos.

Dos enunciados sobre os seres gera-se o chamado conceito, unidade de conhecimento referente ao ser percebido, componente essencial do conhecimento a ser representado. O processo de produção dos registros de conhecimento compreende a etapa de representação da coisa ou ser, gerando-se em decorrência um produto final, um conhecimento sobre a coisa [...]. (ALVARENGA, 2003, p. 21)

O produto final é o conhecimento, representado em forma de documento e que podem ser acessados via conceitos. Para tanto, este processo de representação da informação ocorre em dois níveis. Alvarenga (2003), dá a ideia de que a representação, no nível primário, começa no processo cognitivo de classificação das coisas em nossa mente.

Enquanto a representação primária é feita pelo próprio autor da informação, que é baseado no meio em vive, sua cultura, seu conhecimento e sabedoria, ou seja, o próprio documento. Tomando como base Guedes e Dias (2010, p. 41) o nível secundário é realizado “[...] quando os objetos (documentos) produzidos pelo homem se tornam parte de um conjunto documental, como o acervo de uma biblioteca, sendo mais uma vez representados para que possam ser manipulados facilmente”. O nível secundário seria, de fato, a representação dos documentos em um ambiente informacional com o objetivo de recuperação. Em suma, a secundária, também conhecida por representação temática da informação, é realizada por profissionais da informação, como os bibliotecários, que identificam subsídios no nível primário, que por sua vez serão usados na decisão de quais conceitos permitirão ao usuário acessar o documento desejado. Portanto, os dois níveis se complementam.

Na representação secundária, prática essencial nos sistemas de informações documentais, esses mesmos conceitos constantes dos registros primários são sucintamente identificados em seus elementos constitutivos fundamentais, escolhendo-se os pontos de acesso fundamentais que garantem a representação desse conhecimento (documento) para fins de futura recuperação. Neste caso, os conceitos constantes dos documentos, assim como suas superfícies de emergência, constituem-se em insumos para a representação secundária [...]. (ALVARENGA, 2003, p. 22)

No campo da representação temática da informação surgiram várias nomenclaturas para o mesmo ato de representar um documento: análise da informação, descrição de conteúdo, análise documentária, descrição de assunto, representação de conteúdo, representação de assunto ou simplesmente indexação. De acordo com Guedes e Dias (2010, p. 42) este “ato de representar um documento através de um conceito é de mesma natureza nos diferentes contextos - físico ou digital”. Segundo Da Silva e Neves, (2013, p. 32), a temática objetiva-se “[...] ao conteúdo informacional dos documentos e permite à identificação do tema ou do

assunto a que se refere através das ações de indexação, elaboração de resumos, classificação, disseminação, busca e recuperação”. Para Catarino e Souza (2011, p. 84), esta ocupa-se da “[...] atribuição de assuntos aos documentos a partir da classificação bibliográfica, da indexação e da elaboração de resumos, facilitando a recuperação de materiais relevantes que dizem respeito a temas semelhantes”. Como produtos desta representação têm-se os termos de indexação, que tratam a informação explicitada tradicionalmente no suporte físico.

3.2.1 Indexação: uma ação temática tradicional

A indexação pode ser encarada como uma ação que objetiva tratar os conteúdos dos documentos e assim representar os assuntos dos mesmos para aperfeiçoar os processos de busca da recuperação da informação em sistemas de recuperação da informação (SRI). Conforme Lancaster (2004, p. 12), a indexação “[...] é definida como um conjunto de procedimentos com o objetivo de expressar/representar o conteúdo temático de documentos através de linguagens de indexação ou documentárias visando à recuperação posterior”. Parte deste procedimento temático, como exposto anteriormente, é realizada a priori mentalmente, visão explorada por Lima no conceito de indexação.

[A] indexação é o processo intelectual que envolvem atividades cognitivas na compreensão do texto e a composição da representação do documento. Por ser uma atividade intelectual, utiliza especialmente teorias e métodos da psicologia cognitiva e da teoria de soluções de problemas. (LIMA, 2003, p. 104)

Como um dos produtos da representação temática da informação tem-se os termos de indexação, ou descritores, ou ainda pontos de acesso ou indicadores de assuntos. Segundo Lancaster (2004, p. 6), estes “[...] termos são atribuídos ao documento, no processo de indexação e servem como ponto de acesso mediante os quais um item é localizado e recuperado [...]”.

No processo de indexação têm-se duas etapas principais, isto é, a análise conceitual e a tradução. A primeira etapa é definida por Silva e Fujita (2004, p. 149) como a “identificação de conceitos presentes no texto”. A fase inicial da indexação consta de uma leitura dinâmica do documento onde serão selecionados conceitos, definições, funções etc., candidatos a se tornarem termos representativos ao documento. A segunda etapa acontece à tradução dos conceitos em termos de indexação. Assim a norma da *United Nations International Scientific*

Information System (UNISIST) (1981, p. 83) a define como “a representação de conceitos e, por termos de uma linguagem de indexação”, acordada com os objetivos do SRI.

No processo de indexação, o profissional da informação, com o bibliotecário, atribuirá também indícios de sua vivência, ou seja, do seu ambiente social. Desta forma a subjetividade do bibliotecário também influenciará na indexação, tanto na escolha dos termos, quanto, principalmente, na tradução dos conceitos selecionados para a linguagem do SRI.

O leitor-bibliotecário é um sujeito social sensível às influências culturais, sociais e políticas que o cercam, portanto, é difícil uma neutralidade do profissional nesse processo interpretativo e essa interferência pode influenciar na qualidade (precisão) da informação representada, [consequentemente], imprecisão na recuperação. (MOURA, 2006, p. 30)

Existem alguns fatores que influenciam na qualidade da recuperação da informação, como os parâmetros do SRI, os diferentes tipos de indexação e o tipo de linguagem documentária (de indexação) escolhida pela instituição. Neste sentido o indexador necessita compreender os parâmetros do SRI, ou seja, conhecer o sistema e as diretrizes desenvolvidas para este antes de atribuir os termos. Quanto ao segundo parâmetro, na verdade, durante a prática profissional, existem três tipos de indexação, sendo a inicial representada pela orientação de especialistas, ou seja, realizada por meio de bibliotecários, indexadores e editores no tratamento da informação; a segunda representa a indexação orientada pelo autor, isto é, realizada pelo próprio produtor do documento, ao qual o mesmo insere os termos considerados comuns e aceitos pela comunidade conforme a sua subjetividade; e, por fim, o terceiro tipo, que trata da indexação orientada pelo usuário, enquanto ação realizada pelos receptores da informação ocorre uma interação com a comunidade sem a exigência de seguir regras (GUEDES; DIAS, 2010). Referente ao terceiro e último parâmetro, as linguagens de indexação são construídas a partir da linguagem natural para funcionar como instrumento na tomada de decisão na escolha dos termos. A função básica da linguagem de indexação é orientar o usuário a encontrar o assunto por ele investigado, isto é, funciona também como comunicação entre o pesquisador e o SRI. Corrobora Lara (2004, p. 232) ao afirmar que ao “referir-se ao conjunto dos diferentes tipos de instrumentos especializados no tratamento da informação bibliográfica, designa, de modo mais amplo e completo, a linguagem especialmente construída para organizar e facilitar o acesso e a transferência da informação”.

Para a garantia do sucesso dos SRI existem variáveis chamadas de grandezas inversamente proporcionais, valores que permitem avaliar o desempenho da indexação a partir dos resultados de buscas no sistema, evidenciando assim as falhas no momento da recuperação.

Para Da Silva (2011, p. 48), “as grandezas exaustividade e especificidade estão relacionadas com o processo de organização da informação, as grandezas revocação e precisão se relacionam ao processo de recuperação da informação”. A exaustividade refere-se ao número de termos atribuídos a informação. (LANCASTER, 2004). Em outras palavras, quanto mais exaustiva for a indexação, mais termos serão atribuídos no SRI. Para Rubi (2003, p. 85), esta variável é indicada para “[...] bibliotecas de público variado e de diferentes perfis que podem buscar a mesma informação com termos diferentes”. Já o segundo elemento, a precisão, representa a abrangência do assunto sobre a informação indexada, ou seja, “implica o emprego de uma quantidade muito menor de termos, a fim de abranger somente o conteúdo temático do documento de modo bastante completo” (LANCASTER, 2004, p. 34). As grandezas supracitadas, conforme Piedade (1983, p.11), são “os dois fatores mais importantes que governam a revocação e a precisão”. O índice de revocação (*recall*) se refere à abrangência do número de documentos recuperados sobre o total de documentos os mesmos assuntos contidos na base da instituição (SOUZA, 2006). De acordo com Piedade (1983, p. 11), “a revocação é a relação entre os documentos relevantes recuperados e o número total de documentos relevantes sabidamente existentes na coleção”. A segunda variável, chamada de precisão (*precision*), representa a quantidade de documentos recuperados que são efetivamente relevantes para a busca solicitada pelo usuário, isto é, “a precisão mede o sucesso do SRI em não recuperar documentos que não sejam relevantes de acordo com a necessidade de informação” (SOUZA, 2006, p. 164). Complementando o conceito desta variável, Piedade (1983, p. 11) esclarece que “a precisão é a relação entre os documentos relevantes recuperados e o número total de documentos recuperados”. Como se trata de variáveis que intimamente ligadas à satisfação do usuário, Araújo (1979, p. 48) afirma que “a avaliação desses dois parâmetros é feita medindo-se a relevância de cada referência em relação ao interesse do usuário”. Deste modo, afirma Da Silva (2011, p. 48) que “em termos gerais, o aumento da exaustividade aumenta a revocação e diminui a precisão. Por outro lado, o aumento da especificidade diminui a revocação e aumenta a precisão”.

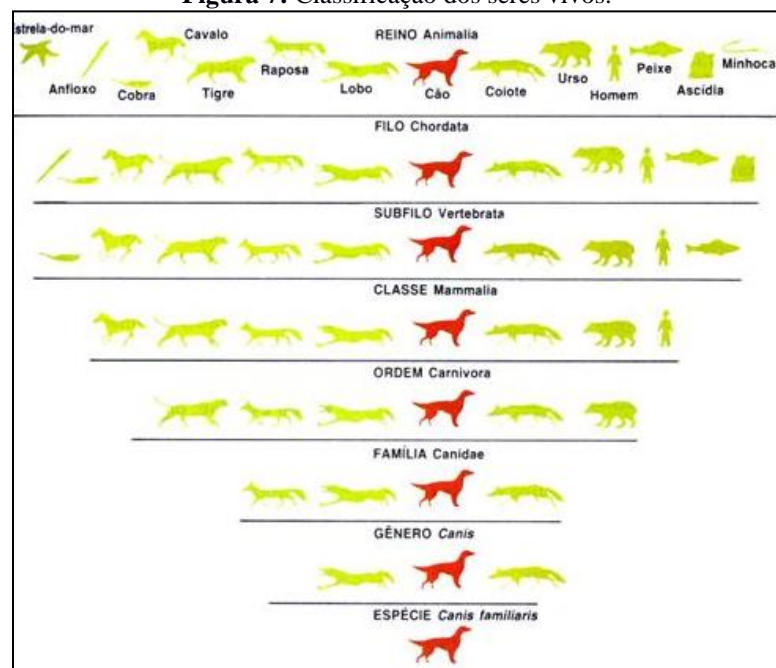
A perspectiva de indexação, apresentada até aqui, é tradicionalmente adotada nos suportes físicos. Já as informações em suportes digitais, advindas das mudanças ocorridas em ambiente da *web*, por exemplo, são representadas tematicamente com outra dinâmica, neste caso, voltadas à participação do usuário neste processo. Diante de uma dinâmica nascida a partir do surgimento da *web 2.0* é possível encontrar na literatura a representação temática abordada como indexação colaborativa enquanto uma ação realizada pelos usuários do ambiente, ou seja, a prática da folksonomia.

3.2.2 Folksonomia: ação temática digital e colaborativa

A indexação orientada pelo usuário pode ser chamada de indexação colaborativa, indexação livre, indexação social etc. Esta forma de representação nasce com a evolução da *web*, considerada a segunda fase do seu percurso histórico.

A fase inicial da *web* tem como característica a comercialização de produtos, dispondo de informações em fluxos unidirecionais, espaço onde o produtor (universidades, empresas e instituições em geral) somente disponibiliza a informação, sem qualquer interação com o usuário, criando um panorama de ambiente estático. Conforme Da Silva (2013), a *web* 1.0 estrutura-se como uma Internet comercial, desenvolvida de forma taxonômica (figura 7) pré-estabelecida (*top-down*), consistindo ao usuário usufruir somente de informações organizadas em arranjos taxonômicos, quase a totalidade em serviços pagos.

Figura 7: Classificação dos seres vivos.



Fonte: Amabis (2007, p. 203).

Tradicionalmente a taxonomia nasceu da classificação dos seres vivos (figura 7), formalizada na botânica. Com base nas falas de Vital e Café (2011, p. 122), a “taxonomia vem do grego *taxis=ordem* e *onoma=nome* e derivou-se de um dos ramos da biologia que trata da classificação lógica e científica dos seres vivos, fruto do trabalho do médico e botânico sueco Carolus Linnaeus (ou Karl Von Linné)”. Esta classificação, dos seres, foi construída de forma hierárquica, dividida por grupos e unidos por características em comum.

No contexto da CI, as taxonomias “são estruturas classificatórias que têm por finalidade servir de instrumento para a organização e recuperação de informação nas empresas”. (CAMPOS; GOMES, 2008, p. 3). Segundo Kato e Silva (2005, p. 1), a taxonomia pode ser definida também como “um vocabulário controlado de uma determinada área do conhecimento, e, acima de tudo, um instrumento ou elemento de estrutura que permite alocar, recuperar e comunicar informações dentro de um sistema sob uma premissa lógica”. Estas definições explicitam as funcionalidades de uma taxonomia em um ambiente *Web* desenvolvida na linguagem de programação (e de marcação) intitulada *HyperText Markup Language* (HTML). Como exemplo temos o antigo diretório do *Yahoo*⁹ (figura 8), que era um índice composto por termos na forma de *hyperlinks*, divididos por categorias e subcategorias. Vale salientar que neste sistema o usuário buscava por assuntos e não por palavras-chaves.

Figura 8: Antigo diretório do *Yahoo*.



Fonte: Google Discovery, 2008.

Com os avanços da tecnologia digital, a *web* foi sendo alterada (reestruturada), quando por volta de 2004 o termo *web 2.0* surgiu em uma sessão de “*brainstorming*” da *O’Reilly Media*¹⁰, editora dirigida por Tim O’Reilly e que atua na área de informática, publicando uma série de conferências feitas pela própria empresa. A segunda fase a *web* “[...] caracteriza-se por

⁹ Website: <https://br.yahoo.com/>.

¹⁰ Website: <http://www.oreilly.com/pub/a/web2/archive/what-is-web-20.html>.

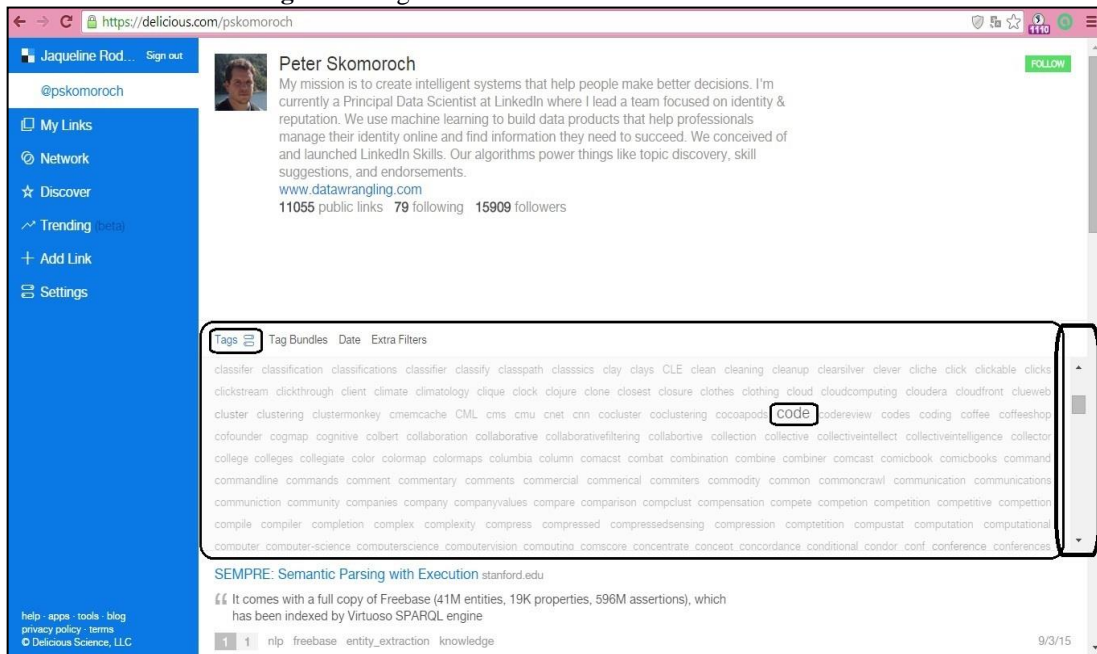
potencializar as formas de publicação, compartilhamento e organização de informações, além de ampliar os espaços para a interação entre os envolvidos nesse processo” (ARAÚJO, 2007, p. 164). Esta geração da *web*, onde o usuário participa ativamente na produção e compartilhamento de informação, “[...] refere-se não apenas a uma combinação de técnicas informáticas, mas também a um determinado período tecnológico, a um conjunto de novas estratégias mercadológicas e processos de comunicação mediados pelo computador” (PRIMO, 2006, p. 1). Complementando, Blattmann e Silva afirmam (2007, p. 197) que a *web* 2.0 descaracterizou a versão 1.0 por propiciar o “[...] nível de interação onde as pessoas poderiam colaborar para a qualidade do conteúdo disponível, produzindo, classificando e reformulando o que já está disponível”.

A *web* 2.0 surgiu como uma plataforma onde as pessoas podem criar e divulgar suas próprias informações e ainda agregar valor a outras informações, como afirmado anteriormente. Nesta fase são os usuários quem modelam a *web* a partir do que fazem e buscam, deflagrando assim a visão *bottom-up*, isto é, as diretrizes partem de baixo para cima, do usuário para o proprietário/desenvolver do ambiente. O sentido vertical de produção e disseminação de informação faz com que o usuário se torne parte essencial para o desenvolvimento do ambiente, onde as características comuns aproximam as pessoas.

Alguns recursos desta fase demonstram, em suas características, as potencialidades do usuário como, por exemplo, o ambiente *Delicious*¹¹, que permite a participação coletiva dos usuários ao colocarem seu *social bookmark* (marcador social) como forma de indexar, armazenar, classificar e compartilhar enlaces na Internet, propiciando o encontro de interesses comuns entre pessoas com o mesmo marcador (termo/*tag*).

¹¹ Website: <https://delicious.com>.

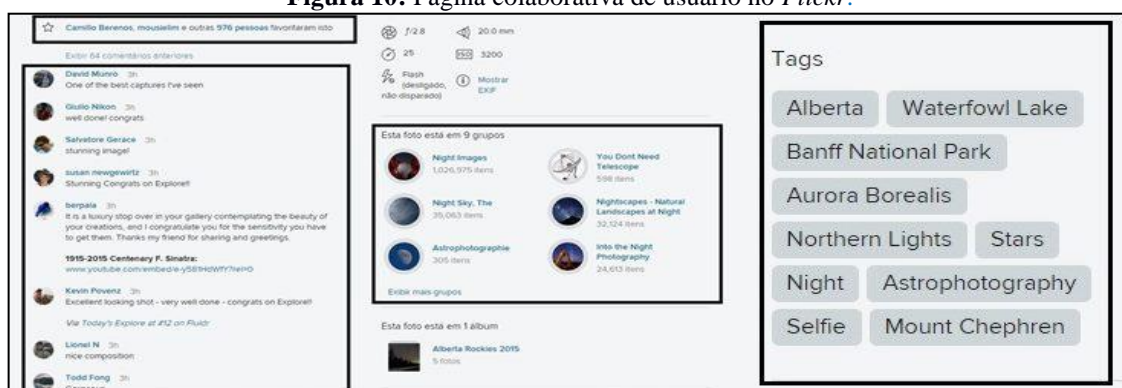
Figura 9: Página colaborativa de usuários no *Delicious*.



Fonte: *Delicious* (2015).

A ferramenta *Flickr*¹² (figura 10), enquanto uma rede social imagética que permite ao usuário inserir imagens na *web*, organizar suas fotos por categorias e grupos segundo *tags* criadas pelo próprio, além da adição de comentários. Este recurso possui a “[...] simplicidade, compartilhamento, atualização constante e participação do usuário que são características que permeiam os recursos da *Web 2.0*” (RUFINO, JESUS; DA SILVA, 2013, p. 5).

Figura 10: Página colaborativa de usuário no *Flickr*.

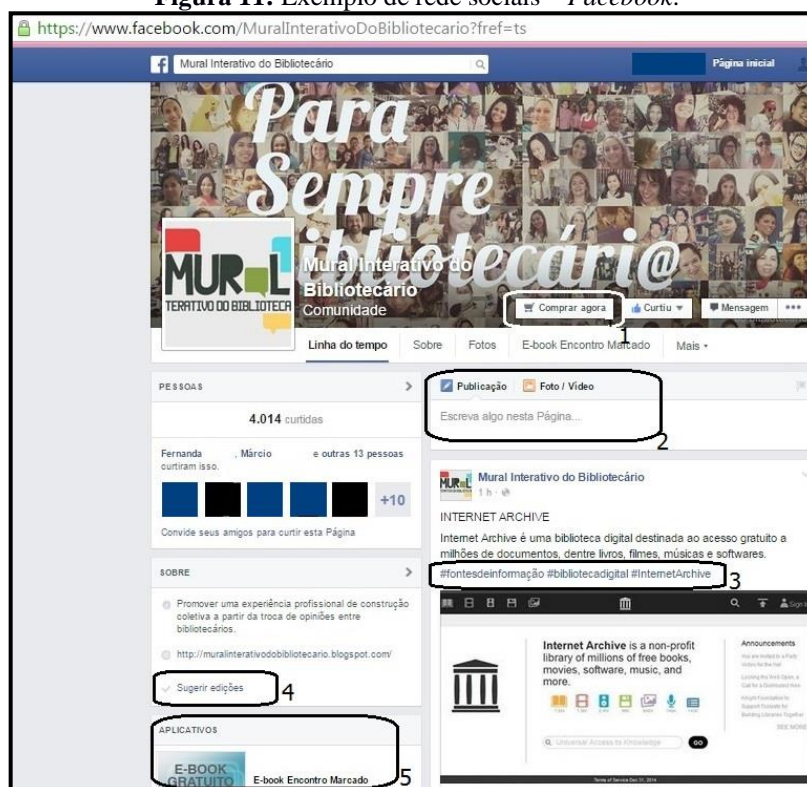


Fonte: *Flickr* (2014).

¹² Website: <https://www.flickr.com/>.

Outra característica da *web 2.0* são as ferramentas de redes sociais como o *Facebook*¹³, que possui diversos recursos para a interação com outros usuários, como *fanpages*, marcação de eventos, atribuição de *hashtags* para organização de textos, imagens e vídeos, compartilhamento de conteúdos etc. A figura 11 ilustra vários recursos além dos supracitados como um *link* para a loja virtual da *fanpage* (1), a publicação de fotos e vídeos (2), a inserção de *tags* para a organização de assuntos (3), a possibilidade de sugerir edições de publicação (4) e a junção da rede social com outros aplicativos que enriquecem o conteúdo do ambiente (5).

Figura 11: Exemplo de rede sociais – Facebook.

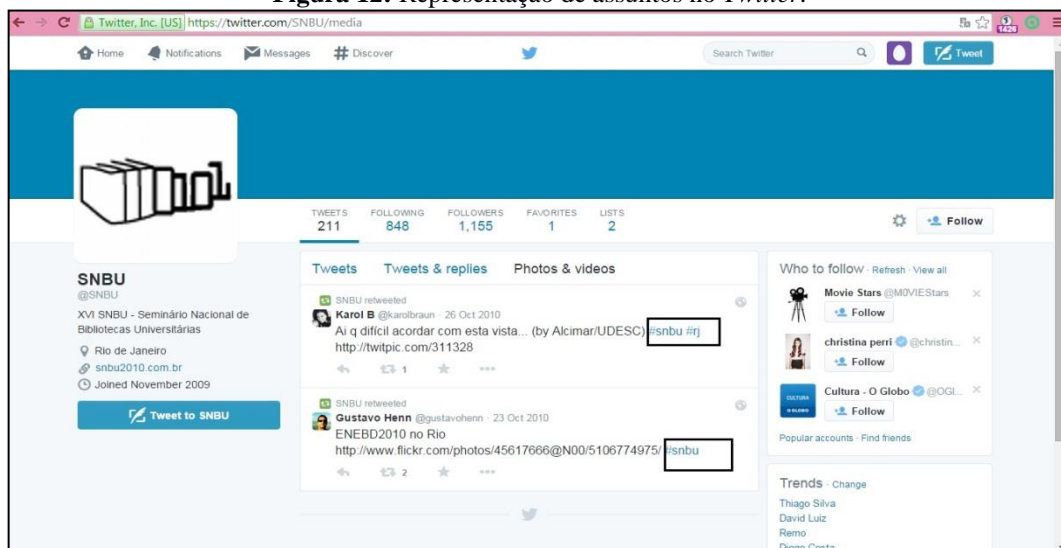


Fonte: Facebook (2014).

O *Twitter*¹⁴ (figura 12) também é um recurso tradicional da *web 2.0*. Trata-se de um microblog que permite o envio de mensagens curtas em 140 caracteres e integração com outras redes sociais, além de ter como destaque o *trending topics*, funcionalidade que mostra as notícias mais comentadas em determinado período na Internet, através da atribuição de *hashtags* para este ranqueamento.

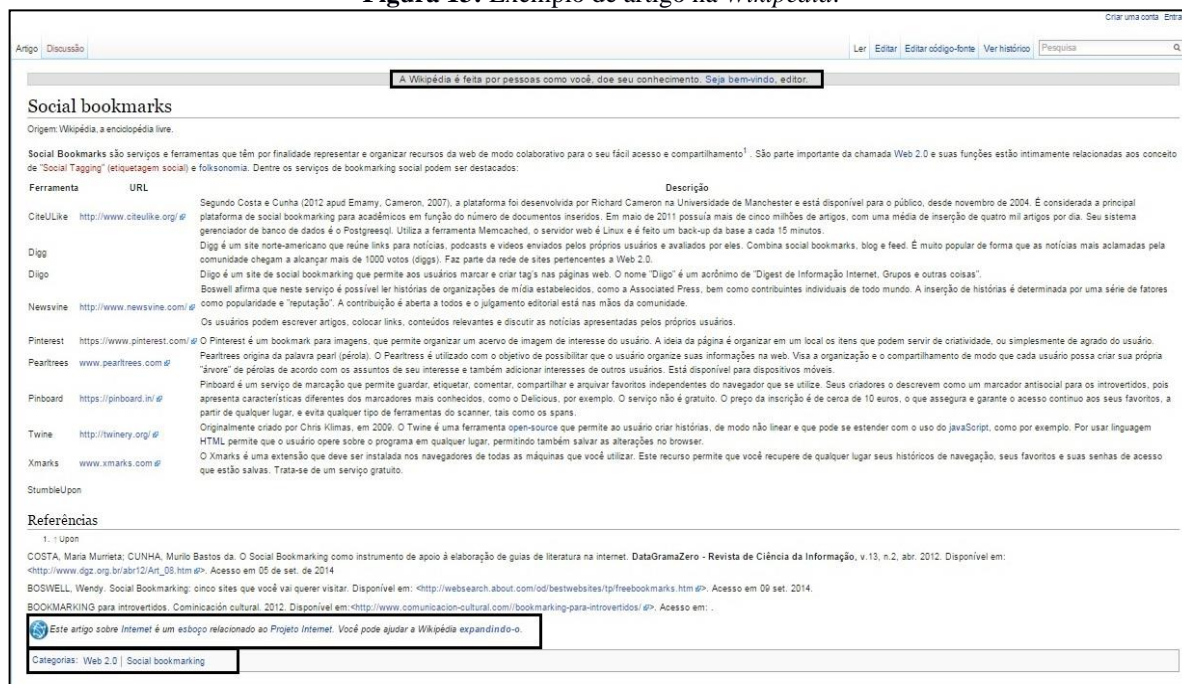
¹³ Website: <https://www.facebook.com/>.

¹⁴ Website: <https://twitter.com/>.

Figura 12: Representação de assuntos no *Twitter*.

Fonte: Twitter (2015).

Outro tipo de recurso da fase da *web* em discussão refere-se a criação e atualização participativa de conteúdos. Como exemplo, temos os ambientes *Wikis*, muito utilizado na criação de enciclopédias livres, sendo a *Wikipédia*¹⁵ a mais conhecida. Esta *Wiki* permite a inserção livre de textos, assim como editar os que já estão presentes no ambiente. Termos podem se transformar em *tags*, que por sua vez estarão relacionados a outros documentos que contenham a mesma *tag*, ambos no formato de *hiperlinks*, como selecionado na figura 13.

Figura 13: Exemplo de artigo na *Wikipédia*.

Fonte: Wikipédia, 2015.

¹⁵ Website: <http://pt.wikipedia.org/>.

As ferramentas supracitadas abrem um panorama para a colaboração do usuário em indexar a informação nos ambientes digitais. As resultantes desta representação podem ser usadas pelos bibliotecários na definição de palavras-chave nas obras presentes nos acervos das bibliotecas, por exemplo. Neste caso, a indexação social não deve ser vista como uma forma de exclusão do tratamento temático da informação tradicional, mas uma ação complementar, agregando valor a linguagem dos usuários conforme características culturais, sociais e temporais. Para tanto ocorre o “tagueamento”, ou seja, a própria indexação social que acrescenta a linguagem natural ao ambiente assim como esclarece Kato e Silva (2010):

O usuário organiza as informações armazenadas (um texto, uma foto, um e-mail, uma página na internet, um filme, etc.) com uma série de palavras-chave (tags) que considera relevantes para descrever e recuperar o conteúdo que está sendo guardado, sem o envolvimento de um profissional especializado e de um vocabulário controlado.

O ato de “taguear” ocorre a partir da categorização de “[...] qualquer palavra que defina a relação entre o recurso on-line e um conceito que está na mente do usuário”. (GUY; TONKIN, 2006, p. 12, tradução nossa) Nesse sentido, as tags são palavras-chaves atribuídas livremente pelo usuário, independentemente de quantidade ou qualidade. Guedes e Dias (2010, p. 47) complementam as potencialidades trazidas pelo “tagueamento” enquanto contribuidores e/ou formalizadores da inteligência coletiva na *web*:

Uma etiqueta (tag) criada por um usuário pode ser utilizada por todos os outros, assim como um item classificado por um usuário também é livre para ser classificado por quem quiser, criando um ambiente favorável ao desenvolvimento do conhecimento coletivo.

A folksonomia é considerada não somente uma forma de representar um conteúdo na *web*, por meio do tagueamento, mas uma forma de organizar tal conteúdo na forma de hipertexto, fazendo uso não mais exclusivamente da linguagem HTML, mas do *eXtensible Markup Language* (XML), que é uma linguagem de marcação capaz de descrever diversos tipos de conteúdos conforme o ambiente digital e necessidades dos usuários, contribuindo no compartilhamento de informações e na construção de outras linguagens.

O termo folksonomia é formado pelas palavras em inglês *folk* (povo) e *taxonomy* (taxonomia), criada por Thomas Vander Wal no ano de 2004, em associação a um grupo de Arquitetura da Informação (AI) e correspondente a forma de representar e organizar a informação na *web* 2.0.

O resultado da atribuição livre e pessoal de etiquetas a informações ou objetos (qualquer coisa com URL), visando à sua recuperação. A atribuição de etiquetas é feita num ambiente social (compartilhado e aberto a outros). A etiquetagem é feita pelo próprio consumidor da informação. (WAL, 2007, tradução nossa)

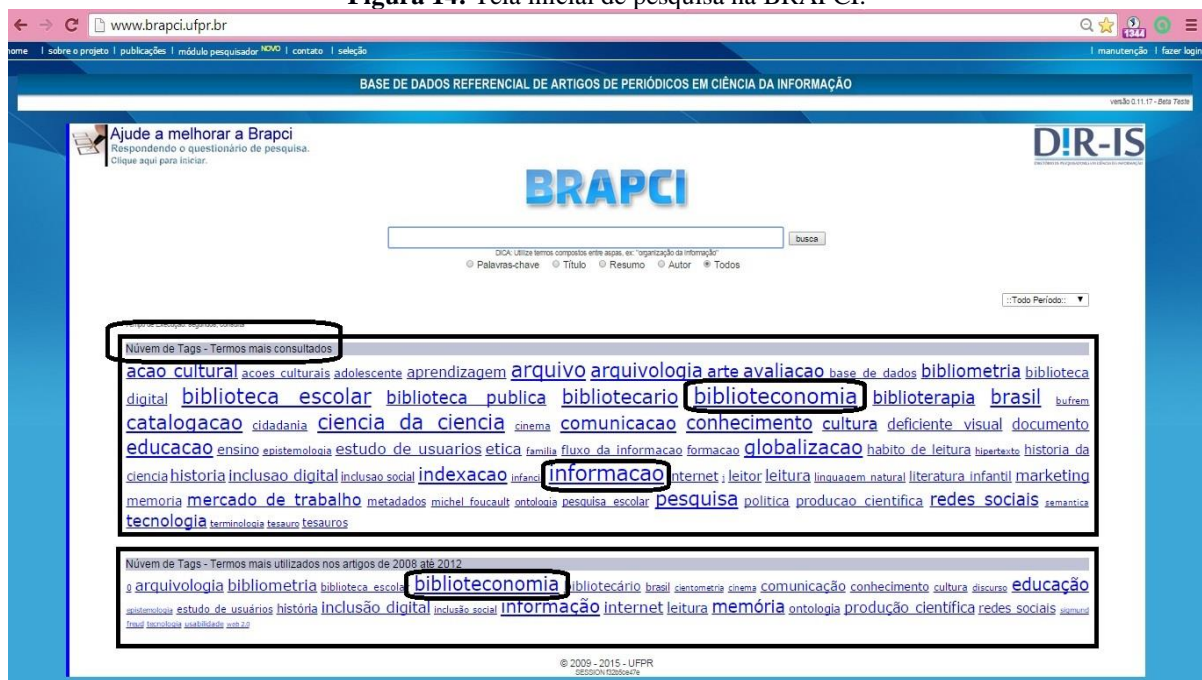
Ao etiquetar suas informações, em um ambiente colaborativo, é promovido o compartilhamento da informação. Além disso, a etiquetagem permite que outros usuários encontrem certas informações pela mesma etiqueta, inclusive, por vezes, consentindo edições como as que comumente ocorrem em ambientes *Wikis*. A etiquetagem é uma ação positiva que permite o aperfeiçoamento da precisão ao recuperar determinado conteúdo conforme rotulações dos próprios usuários e que por vezes se aproximam devido às necessidades em comuns. Neste sentido, os usuários estão familiarizados com os termos a serem adotados em suas buscas.

O fato de a folksonomia ser construída a partir de dados obtidos dos próprios usuários é algo positivo no sentido da garantia do uso, ou seja, o termo usado para representar o documento será o mesmo usado para recuperá-lo posteriormente, por determinado usuário. Neste caso, a garantia de uso do termo na literatura (garantia literária), parece não importar muito, já que a folksonomia é construída a posteriori - não há uma etapa de análise dos documentos do domínio para então se coletar termos e criar posteriormente a base para o esquema de representação. (BRANDT; BRASCHER, 2010, p. 118)

Em suma, a expressão folksonomia refere-se ao processo de representar e organizar a informação como um todo, enquanto o tagueamento pode ser identificado como uma subárea da própria folksonomia. Para Aquino (2007, p. 3), o primeiro “trata-se de um sistema de indexação de informações que permite a adição de *tags* (etiquetas) que descrevem o conteúdo dos documentos armazenados”, enquanto o segundo consiste na escolha de um termo para a representação de determinado assunto/item. Como exemplo citamos a Base de dados Referencial de Artigos de Periódicos em Ciência da Informação (BRAPCI)¹⁶, que aperfeiçoa a recuperação de artigos conforme a geração de uma nuvem de *tags* com os termos mais buscados. Quanto mais a *tag* for usada, maior será o seu destaque na referida nuvem, evidenciada por cores e tamanhos diferenciados em relação as demais. Conforme a figura 14 é possível perceber uma nuvem de *tags* com os termos mais utilizados de 2008 até 2012, período em que foi implementada esta forma de organização e representação no ambiente.

¹⁶ Website: <http://www.brapci.ufpr.br/>.

Figura 14: Tela inicial de pesquisa na BRAPCI.



Fonte: BRAPCI, 2015.

A folksonomia representa um fenômeno realizado pelo usuário ao utilizar sua linguagem natural na representação e organização da informação, locução diretamente relacionada ao cognitivo do mesmo. Brascher e Brandt (2010) afirmam que nas abordagens cognitivas e orientadas a usuários, a unidade organizável é a estrutura cognitiva. Nessa perspectiva, a folksonomia poderia certamente ser considerada uma estrutura de representação do conhecimento. As autoras defendem o fato de que a folksonomia ajuda fortemente na recuperação da informação, indo além do conceito tradicional ao trazer o conceito de cognitivismo à discussão sobre representação no ambiente *web*:

representar o mundo cognitivo do usuário é algo que nenhum outro tipo de representação pode fazer tão bem quanto a folksonomia - ninguém melhor que o próprio usuário para representar seu mundo cognitivo. É sabido que um esquema de representação do conhecimento tão mais preciso será quanto mais restrito e específico for o domínio representado. Levando-se essa especificação ao extremo, deve-se chegar ao usuário, a uma única pessoa - seu mundo cognitivo é certamente o extremo de uma especificação de conhecimento. Sendo assim, a representação do conhecimento gerada numa folksonomia teria uma grande vantagem. (BRANDT; BRASCHER, 2010, p. 119)

Neste sentido faz-se necessário estudar a forma de representação supracitada por meio da identificação de padrões de uso, pois o usuário, ao representar algo conforme seu cognitivo, teoricamente, garantirá a linguagem do público do ambiente. A iniciativa de identificar padrões

permite aos desenvolvedores e proprietários do ambiente *web* “comparar padrões de uso individual e coletivo das *tags*. Uma suposição é que há diferença, ou seja, *tags* usadas por várias pessoas (coletivamente) seguem um padrão diferente das *tags* individuais (usadas por somente uma pessoa)” (GOUVEA; LOH, 2007, p. 3). Complementando sobre o uso da técnica de identificação de padrões na folksonomia, Kato e Silva (2005) afirmam que:

Esses padrões podem ser usados para identificar o surgimento de grupos de interesse em torno de assuntos e temas de maior afinidade, assim como o mapeamento de como os usuários coletam, organizam, compartilham e reutilizam as informações presentes no ambiente em uso.

A identificação de padrões nos permite vislumbrar um ambiente digital organizado de forma híbrida, ou seja, a folksonomia complementando e orientando a taxonomia, assim como a segunda pode complementar a primeira. Conforme Kato e Silva (2005), identificação de padrões pode ajudar “os gestores de bases de informação agir de forma pró-ativa, oferecendo subsídios para as comunidades emergentes, remodelando a navegação, criando novas formas de classificação, otimizando os mecanismos de busca, entre outras ações” no ambiente. Neste caso, em um ambiente híbrido, a recuperação da informação pode ser obtida, tanto pela indexação tradicional, feita pelos bibliotecários, por exemplo, como também pela indexação livre, desta vez, realizada pelos usuários do ambiente.

Diante do contexto apresentado é possível compreender que a folksonomia vai além da rotulação, ou seja, do tagueamento de termos. Neste panorama reforçamos a identificação de padrões, ação que nos permite conhecer mais sobre os gostos, preferências e ações dos usuários no sentido de quantificar, analisar e tomar decisões frente linguagem natural inerente aos ambientes colaborativos. Enxergar a folksonomia desta forma é vislumbrá-la como uma forma de representação do conhecimento da seguinte maneira:

A representação do conhecimento fica desta forma, atomizada, o que suscita a questão da utilidade dessa representação. Além disso, a grande vantagem do mundo cognitivo do usuário é diluída à medida que mais e mais usuários colaboram com suas *tags* em um serviço que vai gerar uma folksonomia. (BRANDT; BRASCHER, 2010, p. 119)

Em suma, a *web* 2.0 é um espaço onde as pessoas se aproximam conforme características em comum. Um mesmo documento pode ser acessado, manipulado e encontrado por uma gama de usuários, indiferente de onde esteja localizado, mas que possuem a mesma necessidade de informação. O documento será representado e organizado pela folksonomia,

entretanto, em tempos atuais, não basta apenas tratá-lo, mas compreender os interesses pelos quais os usuários desejam o mesmo. Estudar as ações dos usuários, via o ato de taggear, pode ser realizada pela técnica intitulada identificação de padrões, que por sua vez gerará valores capazes de análise pela mineração de dados. Neste sentido, o bibliotecário, por exemplo, decidirá sobre o uso de *tags* atribuídas/sugeridas pelos usuários, além de recomendar outros documentos conforme as representações adotadas em seu ambiente, especialmente no digital.

3.3 MINERAÇÃO: ação de descoberta para decisão

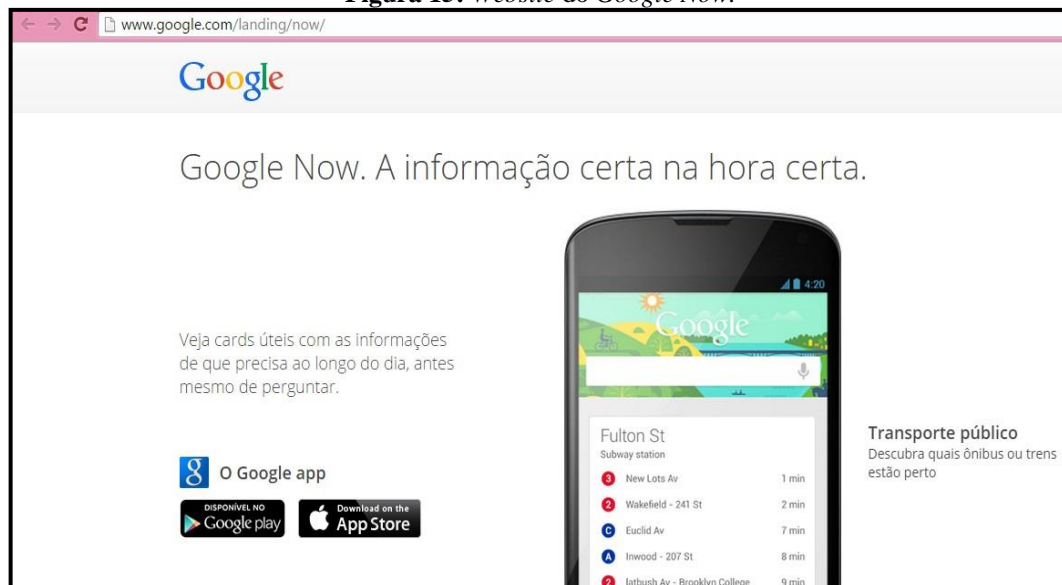
Presenciamos, nas últimas décadas, muitas transformações a partir dos avanços tecnológicos, ocasionando produção em massa de informação. Percebe-se, nesta questão, como a informação tornou-se necessária para o desenvolvimento e crescimento das áreas, entre governamentais, corporativas, científicas e sociais, no caso desta última, deflagrando as chamadas redes sociais, especialmente no ambiente digital. Atualmente a valorização das informações produzidas e disseminadas nas redes sociais é tanta que se tornou um fator estratégico visando a competitividade entre empresas. Neste sentido, armazenar, recuperar, acessar, disponibilizar e agregar valor a essas informações tem sido cada vez mais buscado e questionado, pois apresentarão subsídios para a tomada de decisão pelas organizações.

Entre as ferramentas que levam em consideração as supracitadas ações com fins de análise e tomada de decisão elencamos a mineração, comumente conhecida como mineração de dados (*data mining*). A referida ferramenta de análise capaz de tratar e ler grandes volumes de dados armazenados, especialmente nos tempos atuais com a tecnologia intitulada *big data*, que é um grande banco de dados, alimentado por administradores e usuários, e que apresenta informações em alta velocidade conforme características pré-dispostas pelos que usam um buscador que adota a ferramenta em questão, por exemplo. A ferramenta *big data*, enquanto uma espécie de evolução da tradicional análise de dados, atua na perspectiva dos “5V”, ou seja, velocidade, volume, variedade, veracidade e valor, e pode ser definida, segundo Novo e Neves (2013, p. 32), como “[...] um conjunto de tecnologias e métodos diferenciados para lidar com a análise de dados, permitindo as empresas evoluírem seu modelo analítico para continuar a competir de forma eficiente”. De acordo com Giardelli (2013), a *big data* objetiva analisar interações e transações, entender os dados e tomar a decisão, para gerar mais experiência, produtividade, consumo e novos produtos e serviços.

Como exemplo de ferramentas que analisam interações, entre grandes volumes de dados,

citamos o *Google Now*¹⁷ (figura 15), um aplicativo inteligente que funciona no sistema operacional (SO) *Android*¹⁸, presente em *tablets* e *smartphones*, que por sua vez coletam dados em tempo real, na forma de *big data*, oferecem sugestões, também de forma instantânea, sobre a situação do clima, restaurantes próximos a localização do usuário, dados de voos aéreos, andamento de tráfegos terrestres, transportes públicos na região etc.

Figura 15: Website do Google Now.



Fonte: Google, 2014.

Independente do seguimento da sociedade, o desafio atual que permeia este volume de dados não é unicamente o acesso, mas como armazená-lo e filtrá-lo para a tomada de decisão, que por sua vez subsidiará e/ou iniciará o processo de produção de informação e, conseqüentemente, criará e/ou editará um conhecimento. De acordo com Silva (2004, p. 2), “a informação e o conhecimento obtidos podem ser utilizados para diversas aplicações, que vão do gerenciamento de negócios, controle de produção e análise de mercado ao projeto de engenharia e exploração científica”. Matsuura (2012) afirma que “chegamos uma época em que o processamento, de certa forma, não é tão importante. O que vai definir a informática é como vamos trabalhar com os dados. A gente passou da geração que buscava informação. Agora, é importante filtrar”. Neste contexto encontra-se uma tecnologia de filtragem e gerenciamento de dados denominada *Knowledge Discovery in Database*¹⁹ (KDD) e definida como um método não trivial de identificar em dados padrões que sejam válidos, novos (previamente

¹⁷ Website: <https://www.google.com/landing/now/>.

¹⁸ Website: <https://www.android.com/>.

¹⁹ Processo de descoberta de conhecimento em banco de dados.

desconhecidos), potencialmente úteis e compreensíveis, visando melhorar o entendimento de um problema ou um procedimento de tomada de decisão, ou seja, em dados estruturados realiza-se a análise desses dados à procura de padrões relevantes para determinada questão. O referido processo ocorre nos BD, que segundo Elmasri e Navathe (2005, p. 3), pode ser assim definido:

[...] coleção lógica e coerente de dados relacionados com algum significado inerente [...] é projetado, construído e povoado por dados atendendo a uma proposta específica. Possui um grupo de usuários definidos e algumas aplicações preconcebidas, de acordo com o grupo de usuários.

Ainda definindo BD, Yong (1986, p. 39) afirma que se trata de “[...] um conjunto de dados estruturado de maneira adequada de forma que pode ser utilizado com eficiência por uma diversidade de aplicações dentro de uma organização”.

A estruturação (tratamento) e o armazenamento dos dados em BD permitem a realização de ações intimamente relacionadas a tomada de decisão, entre as quais, conforme Nicholson (2004, p. 260), destacamos a “[...] descoberta e padrões de dados em nível inferior, utilizando ferramentas estatísticas e de inteligência artificial (AI).” É possível identificar no referido contexto a mineração de dados enquanto parte do processo de KDD, já que representa “[...] uma tecnologia que emergiu da intersecção de três áreas: estatística clássica, inteligência artificial e aprendizado de máquina [...]” (SFERRA; CORRÊA, 2003, p. 20).

De forma ampla, há seis termos básicos que são imprescindíveis no processo de KDD, entre dado, padrão, processo, validade, novo e potencialmente útil:

- Dados: conjuntos de fatos;
- Padrão: regra que tenha sentido com o contexto dos dados;
- Processo: sequência de vários passos;
- Validade: verificação da descoberta de padrões, ou seja, do conhecimento obtido;
- Novo: descoberta de um padrão que antes era desconhecido;
- Potencialmente útil: padrões descobertos que são relevantes para a instituição no seu objetivo.

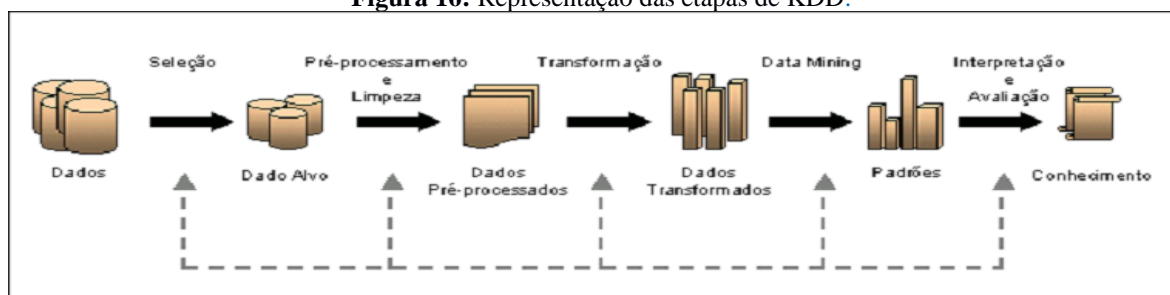
Complementando os supracitados, na prática, Silva (2004, p. 3) defende que o termo mais relevante na descoberta de conhecimento é a compreensão, ao afirmar que:

Um dos objetivos de KDD é tornar padrões compreensíveis para humanos, visando promover uma melhor compreensão dos próprios dados. Embora seja um tanto subjetivo medir compreensibilidade, um dos fatores frequentes é a medida de simplicidade. O fator compreensão dos dados está relacionado á intuitividade da representação destes, bem como da granularidade alta o suficiente para que estes sejam compreendidos. Por exemplo, o \log^{20} de um servidor não é uma representação compreensível; já fatos estatísticos extraídos deste log, tais como totais de acesso ou classificação dos acessos realizados, fornecem informação num formato mais intuitivo e de granularidade humanamente compreensível.

O processo de KDD possui sete etapas e para o sucesso em cada fase é necessário o domínio pleno da ferramenta de mineração, pois este é interativo e iterativo, o usuário pode tomar decisões em qualquer etapa do processo, até mesmo voltar à alguma etapa. Por isto, devem-se ter bem compreendidos os objetivos finais para a melhor obtenção dos resultados.

- Etapa 1: consiste na definição dos dados a serem selecionados;
- Etapa 2: consiste na seleção das variáveis conforme o interesse de aplicação da mineração dos dados;
- Etapa 3: conhecida como pré-processamento dos dados esta etapa consiste na verificação e limpeza dos dados, a fim assegurar a qualidade dos dados;
- Etapa 4: consiste na conversão dos dados para semanticamente ser legível para os algoritmos de mineração;
- Etapa 5: consiste na efetiva busca por padrões, através dos métodos e variáveis selecionadas, resultante que depende diretamente das anteriores;
- Etapa 6: consiste na interpretação dos padrões representados a partir da etapa anterior;
- Etapa 7: consiste no resultado obtido, formalizando determinado conhecimento adquirido e/ou editado e oferecendo subsídios à tomada de decisão.

Figura 16: Representação das etapas de KDD.



Fonte: Figueira (1998, p. 8).

²⁰ Os logs são registros de atividades gerados por programas de computador. (UOL, 2014).

Entre as etapas anteriores, é exatamente a quinta que consiste na mineração de dados, tecnologia que surgiu na área empresarial, especialmente nos Estados Unidos da América (EUA), por volta dos anos 1990, quando os repositórios de dados popularizaram-se, armazenando grandes volumes de dados, que por sua vez eram ferramentas para o desenvolvimento de estratégias competitivas.

Data Mining, ou Mineração de Dados, pode ser entendida como o processo de extração de informações, sem conhecimento prévio, de um grande banco de dados e seu uso para tomada de decisões. É uma metodologia aplicada em diversas áreas que usam o conhecimento, como empresas, indústrias e instituições de pesquisa. Data Mining define o processo automatizado de captura e análise de grandes conjuntos de dados para extrair um significado, sendo usado tanto para descrever características do passado como para prever tendências para o futuro. (SFERRA; CORRÊA, 2003, p. 20)

Complementando o conceito de mineração de dados, Berry e Linoff (2004) afirmam que é uma tecnologia de exploração e análise de grandes volumes de dados e que podem ser realizadas de formas automáticas e semiautomáticas para o descobrimento de padrões e/ou regras. Para tanto, Nicholson (2004, p. 260) cita dois tipos de tarefas desempenhadas para a realização da mineração de dados: a descrição e prospecção.

Na descrição, trata-se de compreender dados passados e presentes. Os padrões descobertos são utilizados para formar “agrupamentos de afinidades” e variáveis comuns a diferentes usuários ou agrupamentos estatísticos com determinadas características. A prospecção, por outro lado, pode ser destinada a fundamentar afirmações sobre o que é conhecido. Pode ser destinada a fundamentar afirmações sobre o que é desconhecido, baseadas sobre o que é conhecido. Pode ser utilizada para prever o futuro ou para fazer afirmações sobre o presente.

A função de descrição está exemplificada na dissertação de Gonçalves (2001) ao apresentar-nos um estudo de caso que utilizou três bases de dados de diferentes supermercados e com contextos diferentes. Contendo dados passados e presentes, os supermercados armazenavam seus dados nestas bases, referentes aos consumidores, sobre o perfil, consumo, datas, horários etc. Com os dados analisados verificaram-se quais produtos poderiam ser comercializados em conjuntos e quais produtos poderiam gerar maiores lucros à rede, acarretando uma sugestão de novos *layouts* para a disposição dos produtos e para a apresentação das lojas.

No caso da função de prospecção, que é uma análise para descoberta de informações em dados já conhecidos, temos como exemplo um estudo de identificação do perfil dos usuários de bibliotecas para a personalização de sistemas *web* de recuperação e disseminação de

informações (JESUS; MOSER; OGLIARI, 2004), aplicado na Biblioteca Central da Fundação Universidade Regional de Blumenau (FURB). A prospecção foi utilizada para ajudar no processo de localização de obras depois identificar o usuário que fez a consulta e, por fim, personalizar os resultados da busca conforme o perfil do investigador. Antes da realização da análise, o sistema apresentava uma alta revocação e pouca precisão, métricas comuns em ambientes mais gerais, abrangentes, como é o caso do *Google*, assim foram analisados os dados dos usuários; as obras da biblioteca; a classificação decimal de Dewey (CDD) do assunto principal da obra e as movimentações do cadastro do usuário. Neste caso tem-se a utilização da mineração para o sistema prospectar realidades e subsidiar a tomada de decisão a partir da análise de dados presentes no cadastro pessoal. Para tanto foi utilizado o *software* chamado *STATISTICA*²¹, produzido pela *StatSoft*²², empresa que atua no âmbito estatístico a partir de ferramentas para análise e gestão de BD com técnicas de mineração de dados.

Os casos supracitados representam uma abordagem tradicional da mineração, com enfoque nos dados. Existe também uma análise que leva em consideração os textos presentes nas realidades em estudo. Cada vez mais usado nos espaços da *web*, a mineração de textos, de acordo com Aranha e Passos (2006, p. 2), é “[...] um conjunto de métodos usados para navegar, organizar, achar e descobrir informações em bases textuais”. Na *web*, por exemplo, os textos provenientes dos usuários, em linguagem natural, são analisados para recomendar opções de busca além da realizada pelo próprio, apresentando uma abordagem de mineração de dados na *web* (*web mining*).

Também chamada por mineração de dados textuais ou descoberta de conhecimento de dados textuais, a mineração em questão busca investigar textos a partir de sua semântica ou estatisticamente (ocorrência de palavras), que podem ser utilizadas juntas, ou não, dependendo do objetivo de sua análise. A análise semântica busca o conhecimento a partir do exame dos termos quanto aos seus significados, morfologia, sintática, a própria semântica e a pragmática, e também o contexto que envolve o texto. Por outro lado, a análise estatística quantifica os números de ocorrência dos termos no texto sem se preocupar com as características da análise semântica.

A mineração de textos utiliza-se de várias áreas do conhecimento: a informática, incluindo a recuperação de informação, o aprendizado de máquina e a inteligência computacional; a estatística; a linguística; e a ciência cognitiva. O conjunto destas áreas subsidiará, na mineração, extrair dados; resumir textos; recuperar informações; descobrir

²¹ Website: <http://www.statsoft.com/Products/STATISTICA/Product-Index>.

²² Website: <http://www.statsoft.com/>.

padrões, associações e regras; e analisar, de maneira qualitativa e quantitativa, um documento em texto. Enquanto a mineração tradicional utiliza dados estruturados ou semiestruturados, o estudo de textos é utilizado na análise de dados não estruturados. Desta forma encontram-se nos dados, em formato de textos, o conhecimento não descoberto previamente útil para a tomada de decisão.

Os dados estruturados são aqueles manipulados em BD, pois estão armazenados em estruturas definidas para assim conseguirem se relacionar com outros dados, através dos esquemas. Os dados semiestruturados são aqueles de *hiperlinks*, ou seja, estão em HTML e XML etc. No caso dos dados não estruturados são aqueles de documentos, pois estão em uma linguagem natural. Em suma, todos os tipos de dados elencados permitem a representação de objetos complexos do mundo real (AMO, 2003).

A mineração de textos é principalmente aplicada na área comercial, através da análise de sentimento em pesquisas de opinião, a partir de questionários de respondidos pelos clientes. No âmbito da inteligência competitiva, a mineração de textos é adotada a fim de capturar informações que agreguem valor ao planejamento tático e estratégico de uma empresa; no suporte e atendimento ao usuário; além de usada na análise e extração de informações em contratos. No direito e na medicina, a mineração de textos é usada na geração de resumos e relatórios/prontuários, pois agilizará a indexação e a recuperação destes documentos (CARRILHO JUNIOR, 2003). As duas últimas áreas citadas vão ao encontro da CI, principalmente na representação temática da informação da biblioteconomia, enquanto um campo que trabalha com dados não estruturados, e na linguagem natural, presentes nos documentos e apresentados na forma textual como resumos, informativos, relatos, notas etc.

Neste contexto elencamos o *software* de mineração de textos intitulado *Text Mining Suite*, que é um conjunto de ferramentas integradas para análise quantitativa e qualitativa dos textos. Enquanto a primeira ação descobre do que os textos estão falando, a segunda descobre o quanto fala-se sobre o assunto identificado no resultado da etapa anterior.

Figura 17: Software text mining suite.



Fonte: Intext²³, 2014.

As perspectivas de análise de dados e textos, nos últimos tempos, formam a técnica de mineração de dados na *web*, que utiliza os dados armazenados em um servidor *web*, fazendo uso do histórico das navegações dos usuários. Este método é visto, segundo Machado e Becker (2002, p. 117), como uma "[...] classe de aplicações que visa identificar padrões de acesso através da análise da interação do usuário com a *web*, possibilitando descobrir informações novas e úteis sobre o comportamento de um usuário durante sua navegação". Há três tipos de categorias de mineração de dados na *web*: de conteúdos, de estrutura e de uso.

A mineração de conteúdos busca descobrir e analisar os dados dentro dos documentos, conhecida também por *web content mining*. Esta técnica é definida por Magalhães (2008, p. 7) da seguinte maneira:

A mineração de conteúdo descreve a descoberta de informações úteis de conteúdos, dados e documentos da Web, através da busca automática de informação de pesquisas on-line. Vale salientar que o conteúdo da Web não se constitui apenas de texto ou hipertexto, mas abrange uma ampla variação de tipos de dados, tais como áudio, vídeo, dados simbólicos, metadados e vínculos de hipertexto.

²³ Website: <http://www.intext.com.br/>.

No caso da mineração de estrutura, objetiva descobrir e analisar dados nos códigos de programação e marcação como o HTML entre os documentos da *web*, conhecida por *web structure mining*. Magalhães (2008, p. 5) novamente contribui na apresentação das categorias de mineração da *web* ao afirmar que:

[A mineração de estrutura] procura descobrir um modelo sobre a estrutura de links da Web. O modelo é baseado na topologia de hiperlinks, com ou sem a descrição destes links. Este modelo pode ser usado para categorizar páginas Web e ser útil na geração de informações similares e relacionadas entre diferentes sites.

A mineração de uso analisa e posteriormente descobre informações a partir da interação com o espaço e uso de conteúdos pelos usuários na *web*, conhecida também por *web usage mining*, compreendida da seguinte forma que:

[...] refere-se à descoberta de padrões de acesso através da análise de interação do usuário com páginas Web. Os dados de uso da Web incluem basicamente os dados obtidos através dos registros de acesso aos servidores Web. A mineração do uso da web está focada em técnicas que possam descrever e prever o comportamento do usuário, enquanto esse estiver interagindo com o site. (MAGALHÃES, 2008, p. 6)

Para ilustrar e aperfeiçoar o conhecimento sobre a mineração *web*, apresentamos a figura 18 que demonstra suas três categorias e respectivas ações.

Figura 18: Abordagens da mineração de dados na *web*.

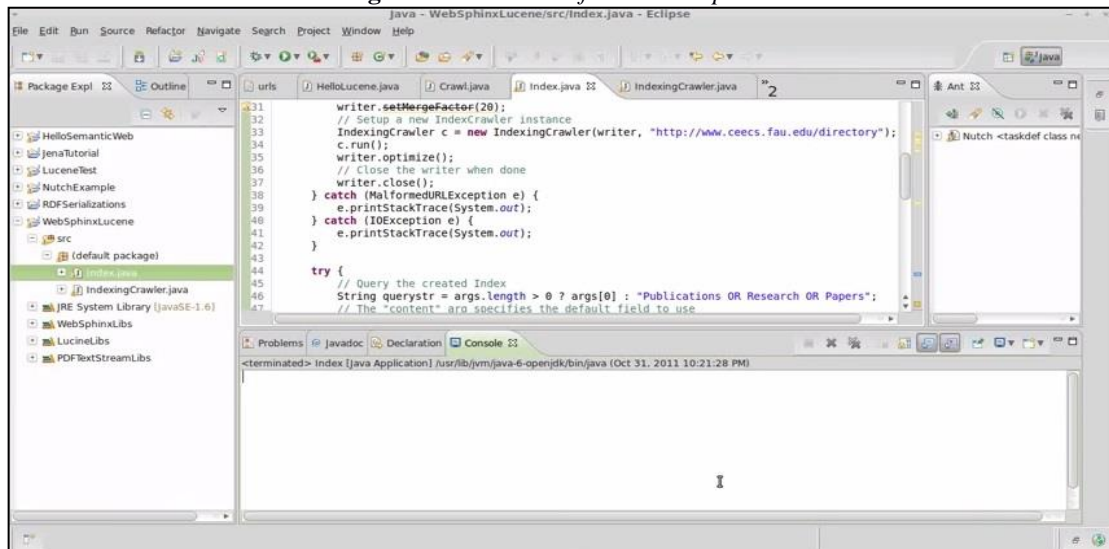


Fonte: Carrilho Junior (2008, p. 2).

Como exemplo de uso da mineração de dados na *web* temos o caso de análise na busca de conhecimento ao verificar as opiniões sobre marcas de *notebooks*. No estudo de Guedes, Afonso e Magalhães (2010), foram pesquisados cerca de 30 ambientes por meio da utilização do buscador *Google*. Em seguida aleatoriamente foram selecionadas sete marcas de *notebooks*.

Após a coleta das opiniões, estas foram organizadas em um arquivo produzido no *software Websphinx*²⁴, sistema que navega e processa páginas da *web* conforme predefinições de busca.

Figura 19: Tela do *software Websphinx*.



Fonte: Websphinx, 2014.

A principal preocupação deste estudo era a extração de elementos textuais, especificamente pequenas expressões que caracterizassem o produto, ou seja, palavras-chaves que expressassem os sentimentos dos usuários, entre positivos e negativos. Os termos que representam as características dos *notebooks* foram analisados conforme a sua frequência de ocorrências. A partir dos resultados foi possível qualificar as marcas dos produtos, encontrando padrões de termos que representassem a opinião dos usuários acerca deles.

Para as futuras tomadas de decisão, os termos extraídos foram estruturados e analisados pelo *software Waikato Environment for Knowledge Analysis (WEKA)*²⁵ (figura 20), sistema desenvolvido na Universidade de *Waikato*, na Nova Zelândia. O *WEKA* é desenvolvido em código *Java*²⁶, disponível na *web* e licenciado sobre o *General Public License (GPU)*²⁷.

²⁴ Website: <http://www.cs.cmu.edu/~rcm/websphinx/>.

²⁵ Website: <http://www.cs.waikato.ac.nz/ml/weka/>.

²⁶ *Java* é uma linguagem de programação e plataforma computacional. (JAVA, 2015).

²⁷ É uma forma de *copyleft* criada para uso em manuais, livros texto ou outros documentos para garantir que qualquer um tem a real liberdade de copiar e redistribuí-los, com ou sem modificações, tanto comercial quanto não-comercialmente. (GNU OPERATING SYSTEM, 2015).

Figura 20: Tela inicial do software WEKA.



Fonte: WEKA, 2014.

A ferramenta contempla uma série de algoritmos de preparação de dados, de aprendizagem de máquina (mineração) e de avaliação de resultados, utilizando técnicas de mineração de dados com fins de descoberta de padrões e assim gerar hipóteses e compreensões, além de fornecer relatórios com dados analíticos e estatísticos.

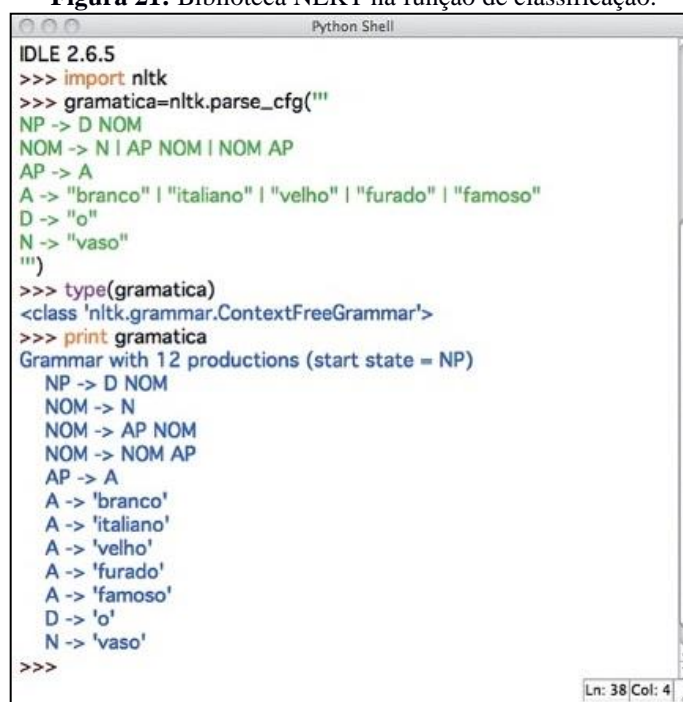
Ainda na *web* existe também a chamada mineração de dados sociais em sistemas de recomendações, que por sua vez são comumente usados em ambiente de comércio eletrônico como a *Amazon*²⁸, ambiente que comercializa livros (impresso e *e-book*), leitores (*e-readers*) etc., é considerada a pioneira no mercado de vendas pela Internet dos EUA. A empresa utiliza métodos de análise das palavras-chaves coletadas nas publicações (postagens) de seus usuários para o funcionamento do próprio sistema de recomendações. A programação deste ambiente funciona com base nas recomendações dos usuários do ambiente ao fornecer uma indicação/comentário de um produto. Com isso o sistema gera a entrada e direciona tal produto para os indivíduos considerados potenciais clientes conforme características que os aproximem, como preferências informadas no cadastro, compras realizadas, navegação no *website* etc.

Como exemplo de estudos sobre a mineração na *web*, especificamente no âmbito das recomendações de produtos, citamos a tese intitulada “Mineração de dados em sistemas de recomendação da *Web*”. Nesta pesquisa, Silva (2011) coletou postagens (*posts*) negativas em relação à opinião de clientes aos produtos da *Amazon* a fim de verificar padrões nestas mensagens. Desta forma foram coletadas as 50 postagens mais recentes relacionadas a entrega dos produtos. Após a coleta foi realizada uma análise gramatical com o objetivo de verificar se realmente as mensagens estavam no contexto de avaliação e assim estabelecidas palavras-chaves segundo critérios pré-estabelecidos. Em seguida cada postagem foi avaliada com base

²⁸ Website: <http://www.amazon.com.br/>.

na frequência com que apareciam as palavras-chaves definidas. O resultado deste caso mostrou que as queixas se concentravam no serviço de entrega da empresa e não nos produtos, resultado que deve ser avaliado pela *Amazon* no sentido de evitar redução nas vendas, especialmente os produtos que possuem alto índice de postagens negativas. Para a análise foi utilizada a biblioteca do *software* de mineração de dados de redes sociais, o *Natural Language Toolkit* (NLTK)²⁹, que processa linguagem natural: “[...] uma plataforma líder para a construção de programas em linguagem de programação Python³⁰ para trabalhar com dados da linguagem humana” (NLKT, 2014, tradução nossa). A função do *software*, de categorização de palavras, foi útil na pesquisa supracitada para a classificação das opiniões como positivas ou negativas.

Figura 21: Biblioteca NLKT na função de classificação.



```

Python Shell
IDLE 2.6.5
>>> import nltk
>>> gramatica=nltk.parse_cfg("""
NP -> D NOM
NOM -> N | AP NOM | NOM AP
AP -> A
A -> "branco" | "italiano" | "velho" | "furado" | "famoso"
D -> "o"
N -> "vaso"
""")
>>> type(gramatica)
<class 'nltk.grammar.ContextFreeGrammar'>
>>> print gramatica
Grammar with 12 productions (start state = NP)
NP -> D NOM
NOM -> N
NOM -> AP NOM
NOM -> NOM AP
AP -> A
A -> 'branco'
A -> 'italiano'
A -> 'velho'
A -> 'furado'
A -> 'famoso'
D -> 'o'
N -> 'vaso'
>>>
Ln: 38 Col: 4

```

Fonte: NLKT, 2014.

Além das ferramentas utilizadas nas pesquisas citadas anteriormente, temos no quadro um (1) um arranjo comparativo entre recursos de *software* de mineração de dados conforme o desenvolvedor, ano, idioma, SO requerido, configuração mínima de *hardware*, preço, tipos de bases de dados que permite a interoperabilidade, regras que o *software* dispõe e os formatos de apresentação dos resultados.

²⁹ Website: <http://www.nltk.org/>.

³⁰ Python é uma linguagem de altíssimo nível, de sintaxe moderna, orientada a objetos, interpretada via *bytecode*, com tipagem forte e dinâmica, modular, multiplataforma, de fácil aprendizado e de implementação livre. (PYTHON BRASIL, 2015).

Quadro 1: Ferramentas de mineração de dados.

Ferramentas	Aira Data Mining®	Xaffinity®	Poly Analyst®	Super Query Discovery Edition®	Super Query Office Edition®	CBA®
Item						
Desenvolvedor	Hycones IT	Exclusive Ore	Mega Puter	Azmy	Azmy	National University of Singapore
Ano	1996	2000	1994	*	*	*
Idioma	Português e Inglês	Inglês	Inglês	Inglês	Inglês	Inglês
Sistema Operacional	Windows	Windows	Windows	Windows	Windows	Windows
Configuração Mínima	Pentium, 32MB Ram, 20 MB HD	Pentium 266, 64 MB Ram	Pentium 266, 32 MB Ram, 20 MB HD	Pentium 266, 32 MB Ram, 20 MB HD	Pentium 266, 32 MB Ram, 20 MB HD	Pentium , 32 MB Ram, 5 MB HD
Preço (US\$)	5.000	5.000	*	449,95	149,95	Livre ⁵
Bases de Dados	Oracle, Sybase, Informix, SQL Server, Paradox e Dbase	SQL Server, MS Access, Oracle, Red Brick e White Cross	MS Excel, ODBC, Oracle, IBM Visual Warehouse, arquivos .csv	SQL Server, MS Access, Oracle e dBase	MS Excel e MS Access	Texto
Regras Geradas	Associativas e hierárquicas	Associativas e sequenciais	Associativas	Associativas	Associativas	Associativas e hierárquicas
Formato dos Resultados	Regras em MS-Word e HTML	Regras em linguagem natural e em colunas	Lista de regras, histograma, gráficos 3D e 2D	Lista de regras e gráfico com os valores mais frequentes	Lista de regras e gráfico com os valores mais frequentes	Lista de regras

Fonte: Gonçalves, 2001.

Nesse contexto podemos perceber a prática da mineração de dados enquanto uma ferramenta capaz de analisar grandes volumes de dados com fins de realizar o processo de KDD. Além da mineração tradicional, existe a realizada em textos, que analisa dados não-estruturados visando a exame da linguagem natural, bastante utilizada em áreas acadêmicas e cada vez mais frequente nos ambientes da *web*. No ambiente digital anteriormente citado, deflagra-se a mineração na *web*, que analisam dados semiestruturados quanto a três aspectos do ambiente conteúdo; estrutura e uso. O primeiro analisa o conteúdo informacional do ambiente digital, o segundo faz a análise da estrutura, ou seja, o(s) código(s) de programação, e o terceiro aspecto refere-se a análise das interações e padrões de acesso dos usuários nos ambientes para aperfeiçoamento da navegação, dos produtos e serviços oferecidos em geral.

A breve discussão apresentada nesta subseção foi necessária, especificamente para o conhecimento de ferramentas que permitam a extração de conhecimento da linguagem natural, pois é onde se descobre a sabedoria dos usuários, certame que se realiza na *web 2.0*, através dos textos e *tags* gerados pelos usuários em espaços como fóruns de discussão, comércio eletrônico e redes sociais etc.

4 METODOLOGIA

Os diversos conhecimentos (popular, religioso, filosófico etc.) produzidos e/ou obtidos pelo homem, durante sua história, permitiram descobrir e compreender os fenômenos da natureza, o que está a sua volta. Para tanto o homem necessitou elaborar e/ou adotar métodos e técnicas, resultando no que chamamos de ciência. Neste sentido, a metodologia é o modo de fazer ciência, ou seja, são procedimentos que o pesquisador deve seguir para realizar sua pesquisa e alimentar a comunicação científica, gerando inovações para sua área de pesquisa. Para Maia (2011), a “metodologia é o conjunto de métodos e técnicas aplicadas para um determinado fim. É o caminho percorrido, a maneira utilizada para atingir o objetivo”, onde se define qual será o tipo de pesquisa, que tamanho será a amostragem, qual será o instrumento de coleta de dados mais adequado e de que forma analisará os dados obtidos.

O presente trabalho consiste em uma pesquisa acerca da representação da informação, em sua vertente temática, confrontando duas abordagens, ou seja, a indexação tradicional oriunda dos espaços físicos e a indexação social realizada em ambientes digitais, neste caso, a folksonomia.

4.1 CARACTERIZAÇÃO DA PESQUISA

De acordo com o problema de pesquisa exige-se a definição de um percurso metodológico constituído de critérios, que devem ser estabelecidos, como o método, natureza, procedimentos técnicos, objetivos e abordagem de coleta de dados, conforme apresentado no quadro dois (2):

Quadro 2: Caracterização da pesquisa.

Natureza	Aplicada
Método	Indutivo
Objetivos	Exploratória
Procedimentos técnicos	Levantamento bibliográfico
Abordagem	Quantitativa e qualitativa
Instrumento de coleta de dados	<i>Software Kitconc 3.0</i> ³¹

Fonte: Elaboração própria com base em Silva e Menezes (2005).

³¹ Para fins dos objetivos da pesquisa não foi necessário utilizar todos os recursos do *software*, ou seja, as abas “colocados” e “n-gramas”. Versão de demonstração (DEMO): <http://www.corpuslg.org/>.

Basicamente existem cinco tipos de métodos: dedutivo, indutivo, hipotético-dedutivo, fenomenológico e dialético. Para a classificação desta pesquisa estipulamos o método indutivo, que de acordo com Prodanov e Freitas (2013, p. 127) ocorre quando “o argumento passa do particular para o geral, uma vez que as generalizações derivam de observações de casos da realidade concreta”. Lakatos e Marconi (2007, p. 86) complementam ao apresentarem a seguinte definição sobre indução:

[...] é um processo mental por intermédio do qual, partindo de dados particulares, suficientemente constatados, infere-se uma verdade geral ou universal, não contida nas partes examinadas. Portanto, o objetivo dos argumentos indutivos é levar a conclusões cujo conteúdo é muito mais amplo do que o das premissas nas quais se basearam.

No que se refere aos objetivos adotamos a pesquisa exploratória, que de acordo com Prodanov e Freitas (2013, p. 52) “tem como finalidade proporcionar mais informações sobre o assunto que vamos investigar, possibilitando sua definição e seu delineamento”.

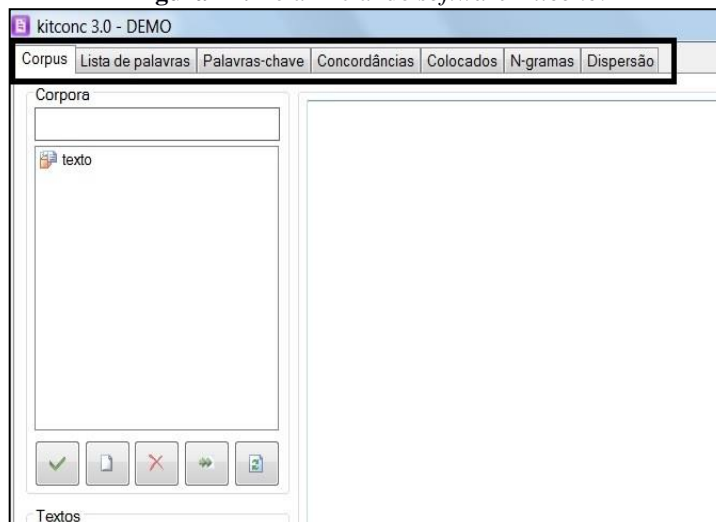
Quanto aos procedimentos técnicos adotamos o levantamento bibliográfico que é “a maneira pela qual obtemos os dados necessários para a elaboração da pesquisa” (PRODANOV; FREITAS, 2013, p. 54) e delineamos a pesquisa com base no problema a ser resolvido durante o percurso metodológico adotado. Conforme Silva e Menezes (2001, p. 21), o levantamento é “[...] elaborado a partir de material já publicado, constituído principalmente de livros, artigos de periódicos e atualmente com material disponibilizado na Internet”.

Para a abordagem do problema da pesquisa utilizamos a quanti-quali. A quantitativa utiliza-se de dados estatísticos, ou seja, toda a parte que envolve dados numéricos. De acordo com Prodanov e Freitas (2013, p. 69), “essa forma de abordagem é empregada em vários tipos de pesquisas, inclusive nas descritivas, principalmente quando buscam a relação causa-efeito entre os fenômenos.” Já a abordagem qualitativa está ligada a interpretação das observações que o pesquisador relatou. Conforme Silva e Menezes (2001, p. 20), “a interpretação dos fenômenos e a atribuição de significados são básicas no processo de pesquisa qualitativa”.

Na coleta de dados, o instrumento definido foi o *software Kitconc* (figura 22), elaborado e desenvolvido por José Lopes Moreira Filho, em sua dissertação de mestrado, para a análise de textos na preparação de aulas. O *software* possui funções de mineração de textos de fácil manuseio e interação contínua, além de ser gratuito (*freeware*), característica adotada como critério de seleção. Outro fator de escolha deve-se a não necessidade em ter conhecimento avançado sobre informática para utilizá-lo, especificamente quanto aos algoritmos utilizados na mineração de textos. De acordo com Silva e Muller (2011, p. 9), o *software* “permite ao

usuário descobrir como palavras ou expressões são utilizados em um texto, apresentando também a frequência de utilização, número de itens e a dispersão dentro do mesmo”.

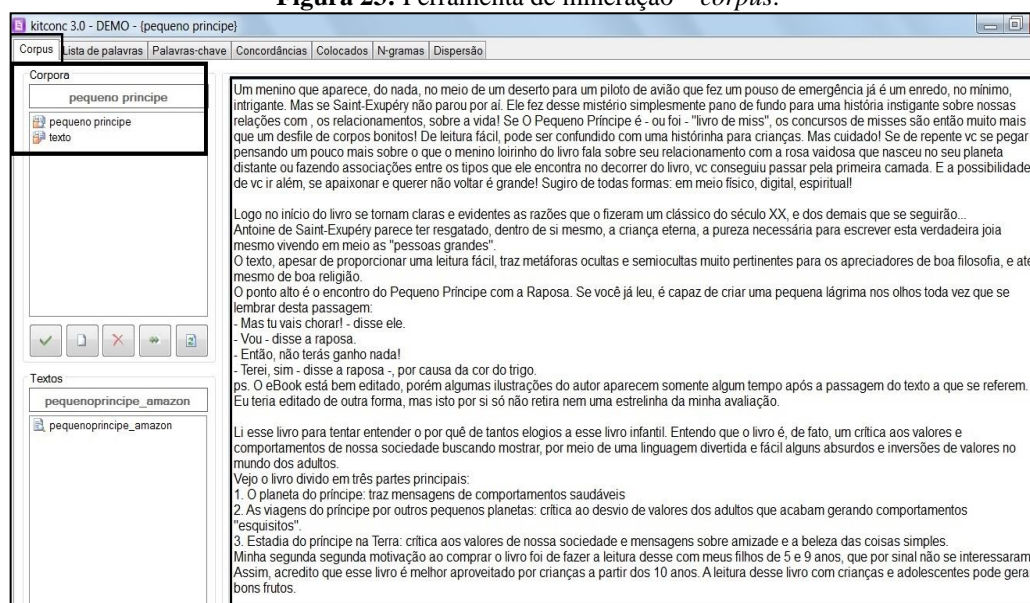
Figura 22: Tela inicial do *software Kitconc*.



Fonte: Kitconc, 2015.

A figura 22 apresenta a tela inicial do *Kitconc*. Na primeira aba temos o “corpus”, que de acordo com Moreira Filho e Zaparolli (2013, p. 2) “[...] deve ser constituído de dados autênticos (não inventados), legíveis por computador e representativos de uma língua ou variedade da língua que se deseja estudar”. Nesta aba o usuário seleciona um arquivo que contém o texto a ser analisado conforme ilustrado na figura 23.

Figura 23: Ferramenta de mineração – *corpus*.



Fonte: Kitconc, 2015.

Na aba “lista de palavras” (figura 24), o *Kitconc* sequencia as palavras quanto à frequência que aparecem no texto. Ao executar esta aba, cada palavra encontrada é adicionada a uma lista. Caso o *software* identifique uma palavra igual (repetida), esta recebe um valor, dada a quantidade de vezes em que foi encontrada. Por outro lado, se a palavra for única, esta é valorada com o número um (1), portanto indicando a frequência única no texto. Os valores também são apresentados em porcentagem (%), seguindo a ordem numeral crescente do item “frequência”.

Figura 24: Ferramenta de mineração – lista de palavras.

N	Palavra	Frequência	%	Marca	Lemas
1	VC	3	00,58		
2	VALORES	4	00,78		
3	UMA	6	01,17		
4	UM	9	01,75		
5	TRAZ	2	00,39		
6	TEXTO	2	00,39		
7	SOCIEDADE	2	00,39		
8	SOBRE	5	00,97		
9	SI	2	00,39		
10	SEU	2	00,39		
11	SEGUNDA	2	00,39		
12	SE	11	02,14		
13	SAINT	2	00,39		
14	RAPOSA	3	00,58		
15	QUE	14	02,72		
16	PRÍNCIPE	5	00,97		
17	POR	8	01,55		
18	PODE	2	00,39		
19	PLANETA	2	00,39		
20	PEQUENO	2	00,39		
21	PASSAGEM	2	00,39		
22	PARA	6	01,17		
23	OU	2	00,39		
24	OS	4	00,78		
25	O	13	02,52		

Fonte: *Kitconc*, 2015.

A aba “palavras-chave” está diretamente relacionada à função mencionada anteriormente, pois seleciona os termos mais frequentes identificados no texto. É a partir da seleção de uma das palavras-chave (figura 25) que serão ativadas as outras abas seguintes, ou seja, “concordância”, “colocados”, “N-gramas” e “dispersão”.

Para a análise e posterior definição das palavras-chaves, o *software Kitconc* oferece a função de “chavicidade”, que sequencia as palavras com base na sua frequência, além de quantificar os termos desta lista em valores de porcentagem.

Figura 25: Ferramenta de mineração – palavras de chaves.

N	Palavra	frequência	%	Chavicidade	Marca
1	LIVRO	11	02,14	69,72	
2	VC	3	00,58	54,05	
3	PRÍNCIPE	5	00,97	49,21	
4	RAPOSA	3	00,58	39,31	
5	COMPORTAMENTOS	3	00,58	35,16	
6	LEITURA	4	00,78	31,45	
7	EXUPÉRY	2	00,39	30,27	
8	VALORES	4	00,78	24,11	
9	EDITADO	2	00,39	20,26	
10	CRÍTICA	3	00,58	19,23	
11	FÁCIL	3	00,58	18,38	
12	SAINT	2	00,39	17,2	
13	MENSAGENS	2	00,39	15,93	
14	ADULTOS	2	00,39	15,61	
15	MEIO	4	00,78	14,75	
16	DESSE	3	00,58	14,11	
17	PLANETA	2	00,39	13,47	
18	MENINO	2	00,39	13,28	
19	CRIANÇAS	3	00,58	12,95	
20	PASSAGEM	2	00,39	11,76	

Fonte: Kitconc, 2015.

Na sequência, a aba concordância (figura 26) apresenta a palavra selecionada no item anterior (palavras-chave), localizada no texto ladeada por 40 caracteres, da direita e da esquerda, por exemplo. Além disso, esta aba pode inferir fatos importantes como o contexto em que está uma palavra selecionada, ou seja, se está acompanhada de um substantivo ou adjetivo, que por sua vez alteram o seu significado naquela oração e podendo ser útil para a reorganização do texto ao alterar as posições das palavras. Também é possível escolher a quantidade de palavras que aparecerá na esquerda ou da direita da palavra selecionada com fins de tornar mais explícito o sentido da palavra selecionada.

Figura 26: Ferramenta de mineração – concordância.

N	Concordância	Palavra	Contexto	Marca	Arquivo
1	tações com , os relacionamentos, sobre a vida! Se O Pequeno	Príncipe	é - ou foi - "livro d...	C:\U...	
2	mesmo de boa religião. O ponto alto é o encontro do Pequeno	Príncipe	com a Raposa. Se você...	C:\U...	
3	o o livro divido em três partes principais: 1. O planeta do	príncipe	: traz mensagens de co...	C:\U...	
4	traz mensagens de comportamentos saudáveis 2. As viagens do	príncipe	por outros pequenos p...	C:\U...	
5	e acabam gerando comportamentos "esquisitos". 3. Estadia do	príncipe	na Terra: crítica aos...	C:\U...	

Fonte: Kitconc, 2015.

A aba “colocados” sequência todas as palavras encontradas conforme a escolha dos usuários, por exemplo, na direita ou na esquerda da palavra selecionada na concordância e tendo a possibilidade de inferir padrões que o significado da palavra constitui no “corpus”, através da frequência obtida. Na figura 27 os caracteres “the” aparecem 72 vezes antes ou depois das palavras selecionadas na aba “concordância”.

Figura 27: Ferramenta de mineração – colocados.

	Palavra	Frequência
1	THE	72
2	A	46
3	AND	33
4	TO	31
5	IN	22
6	OF	22
7	THAT	21
8	I	16
9	BY	15
10	WAS	12
11	FOR	11
12	ON	10
13	NOT	10
14	HE	9
15	BUT	9
16	HIS	9
17	SO	8
18	AT	8
19	VERY	7
20	NO	7

Fonte: Kitconc, 2015.

Sintagmas são identificados na aba “N-gramas” (figura 28), ou seja, obtém-se uma lista de ocorrências de grupos de palavras em sequência e repetidas ao longo do “corpus” com indicação do seu número de repetições (frequência). O usuário pode escolher qual o tamanho dos grupos cuja presença deseja verificar em um intervalo de dois (2) a cinco (5) palavras e assim descobrir possíveis padrões (de repetição).

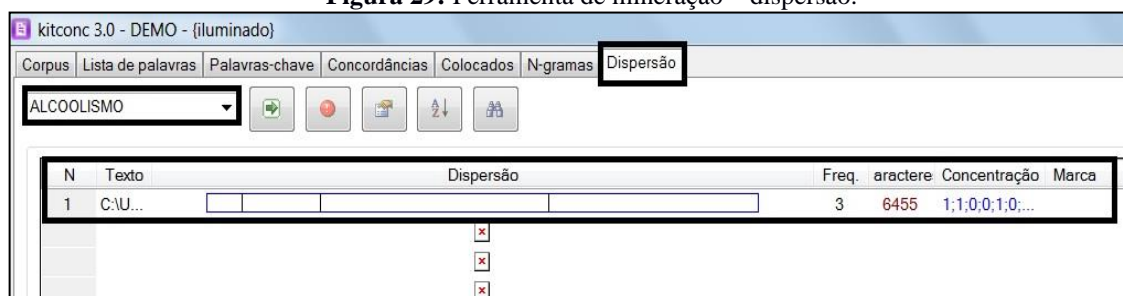
Figura 28: Ferramenta de mineração – N-gramas.

	Pacote	Frequência
1	IT WAS A	9
2	PHILEAS FOGG WAS	8
3	HE WAS A	5
4	THAT HE WAS	5
5	AND THAT WAS	4
6	THE DRIVER WAS	4
7	BUT IT WAS	4
8	BUT THERE WAS	3
9	KNEW WHAT WAS	3
10	WAS GOOD FOR	3
11	BUT HE WAS	3
12	WAS IN HIS	3
13	IT WAS THE	3
14	WAS A MAN	3
15	THAT IT WAS	3
16	WAS THIS WAY	3
17	AS I WAS	3
18	FOR HE WAS	3
19	DOCTOR'S CASE WAS	3
20	WHICH WAS ALWAYS	3

Fonte: Kitconc, 2015.

A aba “dispersão” (figura 29) mostra, através de um gráfico, onde as palavras aparecem no “corpus” do texto com maior frequência. Esta aba é a mais utilizada na correção de textos pela forma imagética (visualização) do “corpus”, porém, por se tratar de uma versão gratuita de demonstração (DEMO), a citada aba não se encontra disponível.

Figura 29: Ferramenta de mineração – dispersão.



Fonte: KitconC, 2015.

No que se refere às etapas da pesquisa, a metodologia seguiu os seguintes passos:

1. Pesquisa bibliográfica a partir do levantamento teórico de autores em livros, artigos, *blogs*, dissertações, tese e materiais didáticos;
2. Escolha do *software* para a mineração de textos;
3. Escolha de dois ambientes para a coleta de dados: um catálogo de biblioteca e um *website* de comércio eletrônico com características da *web 2.0*;
4. Escolha de oito obras aleatórias, presentes nos dois ambientes de pesquisa, constituindo a amostragem, e abordassem assuntos correspondentes a “literatura/ficção”, pois trata-se de uma área do conhecimento presente no sistema de classificação usado na BCE/UnB e que possui considerável quantidade de comentários nos correspondentes livros disponíveis no *website* da *Amazon*;
5. Coleta de dados (textos) nos ambientes da amostragem:
 - a. *Amazon*: três comentários mais recentes sobre a obra e definidos como úteis pelos usuários;
 - b. Catálogo da biblioteca central da UnB (BCE/UnB)³²: termos de indexação coletados no campo assunto.
6. Cadastro dos dados (textos) no *software* de mineração;
7. Interpretação dos dados (textos) analisados pelo *software KitconC*.

4.2 AMBIENTES DA PESQUISA E AMOSTRAGEM

Elencamos, como espaço de coleta de dados, dois espaços da *web*. O primeiro refere-se a um ambiente de comércio eletrônico que utiliza *tags* e comentários para a representação e organização da informação no ambiente, isto é, a *Amazon*, pioneira no ramo de comércio

³² Website: <http://consulta.bce.unb.br/pergamum/biblioteca/>

eletrônico nos (EUA) desde 1994. Sua sede está situada na cidade de Seattle, no Estado de Washington. No Brasil, a *Amazon* iniciou suas vendas no final do ano de 2012.

O *website* da *Amazon* foi escolhido por apresentar interações dos usuários no ambiente acerca de avaliações, compartilhamento e comentários sobre seus produtos. Desta forma a empresa obtém a sabedoria dos usuários, acerca dos produtos disponibilizados, conforme recursos característicos da *web* 2.0 como o *tagueamento* resultante da *folksonomia*, os comentários e avaliações, este último ilustrado na figura 30. No contexto da *web* 2.0 os dados são valorizados por serem inseridos “[...] pelos próprios usuários, integrantes da arquitetura de participação descrita no tópico anterior. Exemplos são os metadados criados e enriquecidos por eles, como as avaliações de produtos em sítios comerciais como Submarino e Amazon [...]” (CAMPOS, 2007, p. 4).

Figura 30: Exemplo de livro no ambiente da *Amazon*.

The screenshot shows the Amazon.com.br product page for 'O Pistoleiro (A Torre Negra Livro 1) [EBook Kindle]' by Stephen King. The page features a navigation bar at the top with the Amazon logo, search bar, and account options. Below the navigation bar is a promotional banner for Kindle devices. The main content area includes the book cover, a star rating of 4.5 from 22 reviews, and pricing information. The Kindle price is R\$ 10,36, while the suggested list price is R\$ 24,90. A 'Compre agora com 1-Click' button is visible on the right. The page also includes a 'Formatos' table and a 'Compre agora com 1-Click' button.

Formatos	Preço da Amazon
eBook Kindle	R\$ 10,36
Capa comum	R\$ 24,20

Fonte: Amazon, 2014.

A *Amazon* permite que os clientes marquem/etiquetem os produtos que desejam comprar, inclusive deixando para o futuro a compra. Basta o produto ser visualizado uma vez, o usuário será informado que existem outros itens semelhantes, que outras pessoas compraram estes produtos junto ao item visualizado. Estes que compraram ou visualizaram o produto podem avaliar e comentar o produto. Os usuários recebem informações também por e-mail cadastrado pelo próprio usuário. A partir dessas informações, especialmente as *tags* atribuídas pelos usuários, a *Amazon* organizará a classificação dos seus produtos, ou seja, a *folksonomia* influenciando a estruturação da taxonomia. Como resultado, a *Amazon* oferece aos seus clientes

um ambiente de organização híbrida e que, conseqüentemente, apresentará maior satisfação nas pesquisas, já que as buscas são filtradas conforme o perfil dos mesmos usuários que participaram da criação de comentários e *tags* anteriormente citadas. Além disso, a *Amazon* também recebe benefícios dos clientes quando os seus produtos são compartilhados em redes sociais, como *Facebook*, *Twitter* e *Pinterest*³³.

O segundo ambiente de pesquisa é o catálogo da BCE (figura 31) da UnB, Instituição criada em 1962 e instalada em um prédio do Ministério da Educação (MEC)³⁴, logo sendo transferida definitivamente para o campus Darcy Ribeiro, situado na Asa Norte, em Brasília, Distrito Federal (DF). A BCE/UnB tem como público alvo os docentes, discentes, pesquisadores e funcionários da Universidade, bem como usuários externos, comumente os que prestam concursos públicos.

Figura 31: Catálogo da BCE/UnB.



Fonte: Website da BCE/UnB, 2014.

A maior demanda da BCE/UnB é suprida por meio do catálogo *online* da rede *Pergamum*³⁵ (figura 32), ou seja, um sistema de automação de bibliotecas (SAB) que gerencia todo o acervo das bibliotecas da rede, proporcionando cadastro, empréstimo, localização, relatórios e outras funções básicas. Os SAB “estão direcionados às exigências propostas pela infraestrutura de *hardware* e *software* que por sinal são bem variadas. Isto é, a partir da

³³ Website: <https://www.pinterest.com/>.

³⁴ Website: <http://www.mec.gov.br/>.

³⁵ Website: <http://www.pergamum.pucpr.br/redepergamum/>.

existência da harmonia entre este adotado atende às necessidades/exigências" da biblioteca (DA SILVA; DIAS, 2010, p. 54).

A BCE/UnB utiliza o *software Pergamum*, que faz parte de um “[...] Sistema Integrado de Bibliotecas, tendo por finalidade melhorar a qualidade global dos serviços dos usuários, promover a cooperação no tratamento da informação e o compartilhamento de recursos de informação” (PERGAMUM, 2015).

Os dois ambientes foram escolhidos para uma comparação entre a indexação tradicional e a indexação social. Para tanto serão utilizados os recursos “corpus”, “palavras-chaves” e “concordância” do *software Kitconc* para que se apresentem como possíveis termos de indexação livre para cada obra escolhida no *website* da *Amazon*. Em contrapartida coletamos os termos com que cada obra escolhida foi indexada no catálogo da BCE/UnB, nos permitindo identificar possíveis semelhanças e diferenças na escolha das palavras-chave entre as duas formas de representação.

A partir da realização das etapas desta pesquisa almejamos verificar quais as contribuições da mineração, especificamente a análise de texto, no processo de representação temática da informação frente ao tradicional modelo de indexação adotado nas bibliotecas. Trata-se aqui de uma comparação discursiva, com fins de proporcionar mais um espaço de atuação do profissional da informação, neste caso, aos bibliotecários. A base para a contribuição da mineração, especificamente em um ambiente da *web 2.0*, deve-se a literatura acerca da tríade dado-informação-conhecimento, que por sua vez acarreta a aplicação do último elemento citado, gerando o que chamamos de sabedoria.

De forma ampla a sabedoria apresenta-se no *website* da *Amazon*, assim como nos demais ambientes de comércio eletrônico, como um resultado da participação do usuário na *web 2.0* a partir da criação, utilização e compartilhamento de informações oriundas de *tags* e comentários dos produtos e serviços, os quais podem contribuir ao conhecimento utilizado pelos bibliotecários na indexação tradicional.

5 ANALISANDO OS DADOS

Para fins da pesquisa foram escolhidas oito obras disponíveis no catálogo da *Amazon* e que possuíssem, no mínimo, três comentários de usuários em cada. Em seguida conferimos se os mesmos livros se encontravam também no catálogo da BCE/UnB. Realizada esta verificação, os três comentários mais recentes e definidos (marcados) como úteis pelos próprios usuários foram salvos em um arquivo no formato texto (txt)³⁶ para cada livro. Para tanto usamos o programa intitulado bloco de notas, um processador de textos do SO *Windows*, enquanto para extrair as palavras-chaves dos arquivos e realizar a mineração de textos utilizamos o *software Kitconc*.

Conforme a figura 32 é possível observar as capas dos quatro primeiros livros selecionados com os seus respectivos títulos, autores e ano de publicação, na sequência.

Figura 32: Obras escolhidas (OE) - parte 1.



Fonte: Amazon, 2015.

- OE1 – Os cinco porquinhos / Agatha Christie / 1984;
- OE2 – O iluminado / Stephen King / 1984;
- OE3 – Mentes perigosas / Ana Beatriz Barbosa Silva / 2003;
- OE4 – O mundo de Sofia / Jostein Gaarder / 1991.

A figura 33 apresenta as capas dos outros quatro livros elencados no catálogo da *Amazon* e descritos pelos títulos, autores e ano de publicação.

³⁶ Foram gerados oito arquivos txt.

Figura 33: Obras escolhidas (OE) – parte 2.



Fonte: Amazon, 2015.

- OE5 – Memórias póstumas de Brás Cubas / Machado de Assis / 1999;
- OE6 – Orgulho e preconceito / Jane Austen / 1997;
- OE7 – O pequeno príncipe / Antoine de Saint-Exupéry / 2003;
- OE8 – Por quem os sinos dobram / Autor: Ernest Hemingway / Ano: 1969.

Representando as ações realizadas em cada livro analisado pelo *software Kitconç*, apresentamos os passos concretizados com a obra “Os cinco porquinhos”. Conforme o critério metodológico adotado, a figura 34 apresenta os textos que foram coletados para análise, especificamente os comentários mais recentes e considerados úteis pelos usuários, resultando a avaliação “muito bom” à obra em questão.

Figura 34: Comentários extraídos da obra “Os cinco porquinhos”.

← → C www.amazon.com.br/Os-cinco-porquinhos-Agatha-Christie-ebook/dp/B00LFTCM18/ref=sr_1_1?ie=UTF8&qid=1426861444&

Avaliações mais úteis

Esta avaliação foi considerada útil por 2 de 2 pessoa(s):

★★★★★ **Diferentes Perspectivas** 25 de setembro de 2014

Por Ielila gonçalves AVALIADOR Nº 1

Formato: eBook Kindle | [Compra verificada](#)

ESSE COMENTÁRIO SE REFERE AO LIVRO CUJO ASIN É B00LFTCM18.

Não é por acaso que "Os Cinco Porquinhos" é considerado um dos melhores livros de Agatha Christie. Publicado em 1941, ele propõe um modelo de ficção policial onde a análise psicológica das personagens é primordial para a elucidação de um assassinato ocorrido no passado.

Portanto, Poirot vai ter que contar unicamente com sua pequeninas células cinzentas, pois nesse caso, não há evidências ou pistas a serem seguidas, apenas lembranças. Aliás, essa história apresenta com clareza a principal diferença entre Poirot e Holmes, enquanto que o excêntrico belga baseia suas deduções no estudo da natureza humana, seu principal concorrente emprega o conhecimento científico.

Os porquinhos do título remontam a uma antiga canção infantil, muito conhecida na Inglaterra, e, curiosamente, estão associados a cinco pessoas diretamente envolvidas na morte do pintor Amyas Crale. São elas: Angela Warren, sua cunhada; Cecília Williams, uma empregada; Elsa Greer, sua amante; Philip e Meredith Blake, amigos e vizinhos.

Contudo, não há dúvida que foi Caroline, cansada das traições do marido, quem o envenenou. Condenada à prisão perpétua, ela faleceu pouco após o julgamento, no entanto, Caria, a filha do casal, não só acredita na inocência da mãe como, às vésperas de se casar, está disposta a reparar a reputação da família.

O assassinato de Amyas é apresentado mediante diferentes perspectivas. Poirot não só entrevista "os porquinhos" como pede a cada um que relate por escrito o que aconteceu. [Saiba mais.](#)

Esta avaliação foi útil para você? Sim Não

Esta avaliação foi considerada útil por 1 de 1 pessoa(s):

★★★★☆ **muito bom** 11 de outubro de 2014

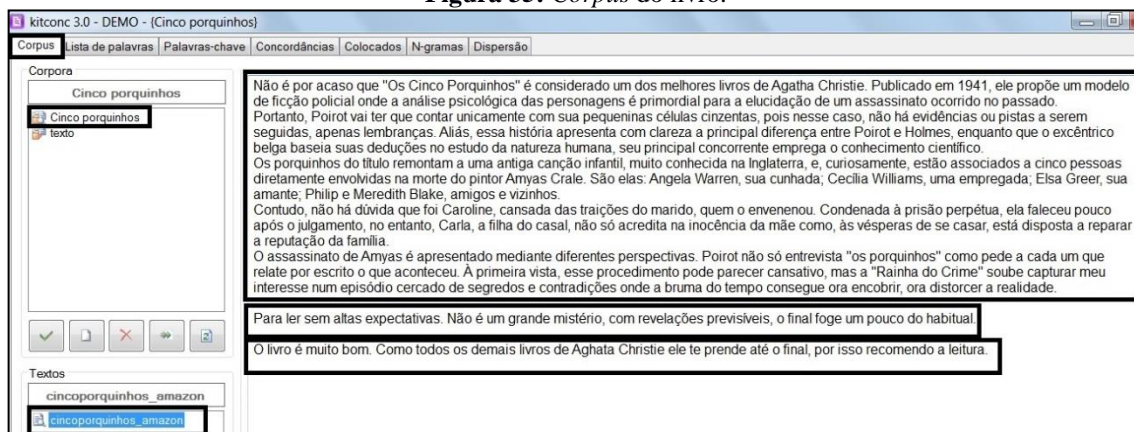
Por lucas

Formato: eBook Kindle | [Compra verificada](#)

Fonte: Amazon, 2015.

Definido o livro no *website* da Amazon, a sequência de ações ocorrida na análise de todos os livros inicia-se pela inserção, na aba “corpus”, dos três comentários selecionados e posteriormente transformados em arquivo txt, salvo com o nome da obra (figura 35).

Figura 35: Corpus do livro.



Fonte: KitconC, 2015.

A partir do “corpus” o *software* lista as palavras presentes no arquivo para que seja iniciada a mineração (figura 36).

Figura 36: Coleta – lista de palavras.

N	Palavra	Frequência	%	Marca	Lemas
1	UMA	2	00,64		
2	UM	6	01,93		
3	SUA	3	00,96		
4	SÓ	2	00,64		
5	QUE	6	01,93		
6	PRINCIPAL	2	00,64		
7	POUCO	2	00,64		
8	PORQUINHOS	3	00,96		
9	POR	3	00,96		
10	POIROT	3	00,96		
11	PARA	2	00,64		
12	OS	4	01,29		
13	ORA	2	00,64		
14	ONDE	2	00,64		
15	O	9	02,89		
16	NO	3	00,96		
17	NÃO	6	01,93		
18	NA	3	00,96		
19	MUITO	2	00,64		
20	LIVROS	2	00,64		
21	HÁ	2	00,64		
22	FINAL	2	00,64		
23	ELE	2	00,64		
24	É	6	01,93		
25	E	5	01,61		

Fonte: KitconC, 2015.

Em seguida, para identificar os termos encontrados nos textos inseridos na etapa anterior, acessamos a aba “lista de palavras” (figura 36), onde foram classificadas todas as palavras quanto a sua frequência de apresentação e um valor em porcentagem (peso) referente à quantidade da citada frequência. A aba em discussão, após coletar e analisar os textos, identificou a presença de 37 palavras distintas, incluindo as *stopwords*³⁷.

Na aba “palavras-chave” foi encontrado um total de 25 palavras (figura 37). Utilizando a classificação de “chavacidade” obtivemos a palavra “poiro” com o índice maior (60,86) entre os termos, ou seja, com a porcentagem de 00,96 e aparecendo três vezes no “corpus”.

Figura 37: Coleta – palavras-chave.

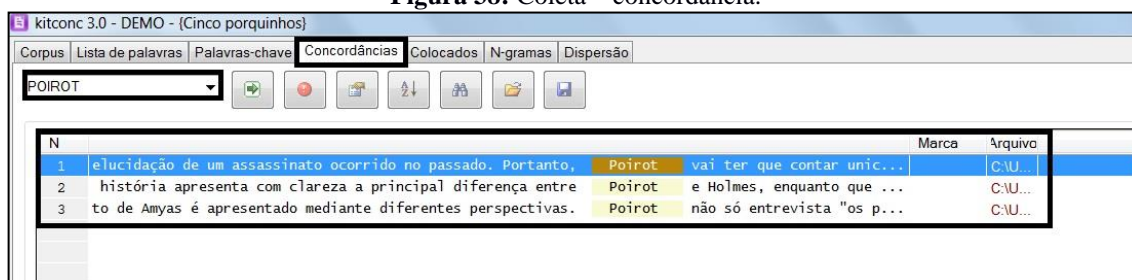
N	Palavra	Frequência	%	Chavacidade	Marca
1	POIRO	3	00,96	60,88	
2	PORQUINHOS	3	00,96	48,96	
3	CHRISTIE	2	00,64	26,92	
4	ORA	2	00,64	16,22	
5	ASSASSINATO	2	00,64	15,74	
6	LIVROS	2	00,64	11,81	
7	PRINCIPAL	2	00,64	8,78	
8	POUCO	2	00,64	7,58	
9	CINCO	2	00,64	7,35	
10	FINAL	2	00,64	5,8	
11	ONDE	2	00,64	5,11	
12	DE	7	02,25	4,88	
13	SUA	3	00,96	4,67	
14	MUITO	2	00,64	4,04	
15	NÃO	6	01,93	3,93	
16	SÓ	2	00,64	3,49	

Fonte: Kitconc, 2015.

Porém foi descoberto que a palavra “poiro” é o nome de um personagem no livro e não poderia ser classificado como um assunto, independentemente de a “chavacidade” ser alta. Esta particularidade foi descoberta a partir da aba concordância (figura 38) ao analisarmos os contextos em que a palavra no “corpus” aparece, podendo contribuir na compreensão dos termos e assim percebendo idiomas diferentes em relação ao apresentado no “corpus”, segundo significados e fonéticas distintas. Foi percebido também que outras palavras com grande “chavacidade”, como “assassinato” e “principal”, são termos que não possuem significado de tematicidade, ou seja, não representam um assunto, portanto não podem ser consideradas palavras-chaves do livro em análise.

³⁷ Em tradução livre significa palavras paradas, ou seja, que não contêm significado para substantivos. Por exemplo, artigos, até, “tbm”, “vc” etc.

Figura 38: Coleta – concordância.



Fonte: Kitconc (2015).

Ao final realizamos um recorte entre as palavras-chave com valor de “chavicidade” de um (1) a oito (8). Em seguida, com fins de definir as palavras-chave de cada livro, selecionados no *website* da *Amazon*, e posterior comparação com as mesmas palavras identificadas no catálogo da BCE/UnB, uma nova filtragem foi feita, ou seja, definimos até cinco palavras-chave com maior concordância ao assunto de cada livro. Todas as palavras supracitadas, bem como os resultados de cada etapa da análise de mineração, no *software Kitconc*, podem ser observadas no quadro três (3).

Quadro 3: Análise de *corpus* de cada livro.

Obras	Palavras-chaves (chavicidade de 1 a 8)	Concordâncias (1 a 5)	Palavras-chaves (1 a 3)	Assuntos
OE1	poirot; porquinhos; christie; ora; assassinato; livros; principal; pouco	poirot; porquinhos; christie; assassinato; principal	assassinato	Assassinato
OE2	overlook; danny; jack; torrance; hotel; wendy; iluminado; king	overlook; torrance; hotel; iluminado; king	-	-
OE3	psicopatas; estes; situações; tipos; gostaria; tivesse; cara; fazer.	psicopatas; cara	psicopatas	Psicopatas
OE4	sofia; postais; filosofia; livro; coragem; hilde; perseverança	filosofia; coragem; perseverança; hilde	filosofia	Filosofia
OE5	machado; assis; idealizações; cubas; brás; lido; narrativa; paradigma.	idealizações; narrativa; paradigma	narrativa	Narrativa
OE6	erros; leitura; achei; vocabulário; comprei; sr; voltada; amor	erros; amor; vocabulário	amor	Amor; Romance
OE7	livro; vc; príncipe; raposa; comportamentos; leitura; exupéry; valores	comportamentos; valores	-	-
OE8	livro; hemingway; ler; romance; seguinte; amor; guerra; dia	romance; amor; guerra	amor; romance; guerra	Romance; Guerra; Amor

Fonte: Da pesquisa, 2015.

Apesar dos termos oriundos dos itens “chavicidade” e “concordância”, as definições das palavras-chave ocorreram conforme uma possível representação temática da obra, listados na coluna “Palavras-chave (1 a 3)” do quadro dois (2). Neste sentido, entre os resultados obtidos pela mineração de texto, realizada pelo software *Kitconc*, inferimos que o resultado da análise da obra OE1 foi positivo ao apresentar o termo “assassinato” como palavra-chave do livro e assim permitir a identificação do assunto, pois trata-se de uma ficção policial. Os livros OE 2 e OE 7 não obtiveram resultados positivos, pois não restaram palavras que pudessem representar o assunto principal da obra. No caso da obra OE4 obtivemos “psicopatas” como palavra-chave, sendo a que possui mais referência ao conteúdo da obra, tratando-se de indivíduos que possuem psicopatologias. Quanto ao livro OE5, apesar do “corpus” ser grande, não obtivemos uma análise positiva quanto ao descobrimento de palavras-chave que representassem o conteúdo do livro, mesmo com as “chavidades” altas. Na análise do “corpus” da obra OE6 encontramos um resultado que não havíamos previsto a priori, pois obtivemos algumas palavras que contêm características (qualificadores) do produto como os termos “erros”, “versão”, “português”, “inglês” e “vocabulário”, mas que não podem ser adotadas como palavras-chaves representativas dos assuntos do livro. Citando novamente o caso do “corpus” da obra OE7, vale salientar que encontramos como palavra-chave o termo “comportamentos”, mas apesar da alta “chavicidade” e possuir concordância em alguns trechos, não pôde ser considerada um assunto do livro pela falta de um complemento que especificasse sobre o que trata a obra em discussão. O conteúdo da OE8 trata-se de um romance que acontece em um período de guerra indo ao encontro da análise realizada, ao qual foi possível verificar as palavras-chave “romance”, “amor” e “guerra” como representantes do assunto da obra. Em suma, a partir das palavras-chave definidas para cada obra foi possível perceber que os materiais analisados tratam de temáticas relacionadas ao assunto literatura em consonância com os resultados dos livros consultados no catálogo da BCE/UnB (quadro 3).

Analisados os livros supracitados e obedecendo aos objetivos da presente pesquisa foi elaborado um quadro comparativo entre as palavras-chave (*tags*) adquiridas no *website* da *Amazon*, representando o ambiente da *web 2.0* e que faz uso da folksonomia, com os descritores apresentados no catálogo da BCE/UnB para representar os livros rotulados pela indexação tradicional, comumente adotada em bibliotecas físicas.

Quadro 4: Comparação de palavras-chave entre os ambientes.

Obras	Amazon	BCE/UnB
OE1	Assassinato	Literatura inglesa; Ficção inglesa
OE2	-	Literatura americana; Ficção americana
OE3	Psicopatas	Psicologia – doentes mentais; Psicologia – comportamento humano; Psicopatologia
OE4	Filosofia	Literatura norueguesa
OE5	Narrativa	Literatura brasileira; Romance
OE6	Amor; Romance	Literatura inglesa; Romance
OE7	-	Literatura francesa; ficção francesa; literatura infanto-juvenil
OE8	Romance; Guerra; Amor	Literatura americana

Fonte: Da pesquisa, 2015.

No quadro quatro (4) podemos perceber que a maioria dos termos elencados pelo *software*, a partir da sabedoria dos usuários nos comentários das obras no ambiente *Amazon*, não são adotadas como palavras-chave representativas dos assuntos de cada obra. Entretanto estas *tags* possuem as qualidades (características) de apresentarem os significados a respeito do conteúdo dos livros, permitirem uma navegação intuitiva entre os materiais segundo as *tags* oriundas da indexação social, oferecerem recomendações de outros livros semelhantes ao pesquisado e subsidiarem os usuários no momento da escolha do livro a partir de avaliações/comentários feitos pela coletividade, diferentemente dos catálogos tradicionais. Neste sentido, apesar de alguns termos não serem definidos como palavras-chave, os mesmos são fundamentais para a descoberta do assunto de cada obra, situação que se deflagra na obra OE8, pois a indexação tradicional apresentou somente um único descritor, “literatura americana”, enquanto que a mineração resultou em duas palavras que representam o conteúdo da respectiva obra, ou seja, aferimos que trata-se de uma obra de literatura americana, especificamente um romance em meio ao decorrer de uma guerra, nos permitindo inferir que o conhecimento sobre o conteúdo da obra pode ser descoberto a partir de palavras identificadas na mineração de texto.

Algumas palavras-chave definidas na mineração de texto se aproximam de termos presentes no catálogo da BCE/UnB, como é o caso da obra OE3, ao verificarmos que, na indexação social, o termo que se aproxima de um assunto é “psicopata”, enquanto que na versão tradicional da indexação, o descritor adotado é “psicopatologia”.

Por outro lado, os descritores das obras selecionadas no catálogo da BCE/UnB aparentemente foram retirados de uma taxonomia e/ou sistema de classificação, como a CDD, instrumentos de representação e organização que apresentam-se como um conjunto de descritores com o objetivo de recuperação, visando somente à informação registrada em seu

ambiente físico, ou seja, representando a classe que indica o endereço das obras na estante e, por vezes, não indicando o assunto com riqueza de detalhes. Nos comentários da obra OE8, por exemplo, foram encontradas, na mineração de textos, palavras-chave que são assuntos do conteúdo do livro, mas que não foram contempladas na indexação do catálogo da BCE/UnB, nos permitindo compreender que consultar unicamente a taxonomias, por vezes, pode ser uma ação limitada no processo de decisão quanto aos termos representativos ao conteúdo da obra.

6 CONSIDERAÇÕES FINAIS

No presente trabalho abordamos a tríade que constitui a sociedade, formalizada pelos elementos dado, informação e conhecimento. O citado ciclo, ao ser aplicado, gera o desenvolvimento, apresentando uma nova estrutura aqui intitulada de ciclo informacional, que por sua vez, ao receber insumos da inteligência coletiva, apresenta a sabedoria, que é o conhecimento explícito em ambientes que estão se inseridos no contexto da *web 2.0*, onde os usuários participam efetivamente na representação e organização da informação, comentando, avaliando, compartilhando e tagueando produtos e serviços.

Os elementos que constituem os caminhos da sociedade, anteriormente elencados, são vislumbrados na representação temática da informação de duas formas: a partir da subjetividade do profissional da informação, como os bibliotecários, ou seja, a indexação tradicional, que é a forma de rotular tematicamente um item físico em um determinado acervo/catálogo, visando a sua posterior recuperação pelo usuário; e a segunda conforme a coletividade no ambientes digitais a partir do tagueamento, ação oriunda da folksonomia e considerada uma representação baseada na linguagem natural do usuário enquanto ação basilar da *web 2.0*.

No cerne do debate entre os elementos que constituem a sociedade da informação e a representação temática da informação, neste trabalho, encontra-se a mineração, especificamente a modalidade de texto, com fins de descobrir de que forma a mineração de dados pode contribuir na representação temática da informação, partindo do pressuposto de que a tecnologia de mineração é utilizada para analisar e descobrir conhecimentos que possam agregar valores estrategicamente para os objetivos estabelecidos, a priori, em grandes quantidades de dados armazenados em BD. Neste sentido, no interesse em confrontar a indexação social, realizada no *website* da *Amazon*, com a indexação tradicional, contemplada no catálogo da BCE/UnB, uma amostragem foi constituída. Percebemos que os resultados obtidos na mineração de texto foram mais fidedignos do que os apresentados no catálogo da BCE/UnB. Direcionando os resultados para as grandezas inversamente proporcionais da indexação, nos parece que os ambientes da *web 2.0* promulgam pela revocação, especialmente pelas iniciativas de navegação e recomendação, enquanto que o catálogo da BCE/UnB adota a precisão como grandeza, ao levar em consideração o tempo do leitor na busca pela informação necessitada, mesmo que as ocorrências não sejam satisfatórias.

A partir da mineração de dados abre-se um panorama para que ambas as formas de indexação aqui abordada, tradicional e social, possam ser usadas de maneira concomitante e não substitutiva especialmente usando a sabedoria dos usuários implícita nos comentários,

“tagueamento” e avaliações para a tomada de decisão quanto aos termos representativos, tematicamente, das obras do acervo. No caso deste trabalho, a mineração de texto contribuiu de forma significativa para a descoberta de padrões para a representação temática da informação a partir dos termos que se apresentaram com frequência e possuíam concordância semântica no texto, tornando-os posteriormente candidatos às palavras-chave que permitirão definir os assuntos das obras.

Apesar de vislumbramos certa contribuição da mineração na representação temática da informação, vale destacar uma dificuldade/limitação pertinente ao uso do *software Kitconc* por não disponibilizar alguns recursos em sua versão gratuita (DEMO), neste caso, a aba “dispersão”, que nos permitiria visualizar imagicamente as posições dos termos com mais frequência ao longo do texto.

Conclui-se que os comentários obtidos na *web 2.0* apresentam o conhecimento dos usuários que usufruíram da obra e explicitaram sua sabedoria em postagens, aqui transformadas em palavras-chave, mostrando que os mesmos podem participar efetivamente na representação temática da informação e contribuir posteriormente no sucesso dos processos de busca, nos permitindo inferir que para uma possível representação fidedigna do conteúdo da obra e melhor recuperação, a indexação social pode subsidiar/complementar a indexação tradicional, ou seja, as *tags* dos usuários, analisadas pela mineração, podem ser comparadas aos descritores definidos pelos bibliotecários para a tomada de decisão quanto aos termos representativos das obras.

Para trabalhos futuros visamos utilizar outros recursos de *software* e em outros ambientes com maiores quantidades, tanto na perspectiva de textos, como também no uso de dados, com fins de apresentar resultados de busca satisfatórios segundo as necessidades dos usuários. Esperamos que este trabalho fomente novas pesquisas e amplie o olhar do bibliotecário frente as inovações tecnológicas, não as estudando como substitutivas, mas sim como um aperfeiçoamento do trabalho já consolidado, como um sistema híbrido que faz uso concomitante das indexações tradicional e social (folksonomia).

REFERÊNCIAS

ALVARENGA, L. Representação do conhecimento na perspectiva da ciência da informação em tempo e espaço digitais. **Encontros Bibli: revista eletrônica de biblioteconomia e ciência da informação**, v. 8, n. 15, p. 18-40, 2003. Disponível em: <<https://periodicos.ufsc.br/index.php/eb/article/viewArticle/97>>. Acesso em: 21 ago. 2014.

AMABIS, J. M. **Biologia 2: Organismos, classificação, estrutura e função nos seres vivos**. São Paulo: Editora: Moderna, 2007.

AMO, S. **Curso de data mining**. 2003. Material didático. Disponível em: <<http://www.deamo.prof.ufu.br/arquivos/Aula2.pdf>>. Acesso em 11 jan. 2015.

AQUINO, M. C. Hipertexto 2.0, folksonomia e memória coletiva: um estudo das tags na organização da web. **E-Compós**, Brasília, DF, v. 9, ago. 2007. Disponível em: <http://www.compos.org.br/files/15ecompos09_MariaClaraAquino.pdf>. Acesso em: 13 ago. 2014.

ARANHA, C.; PASSOS, E. A tecnologia de mineração de textos. **Revista Eletrônica de Sistemas de Informação**, v. 5, n. 2, 2006. Disponível em: <<http://189.16.45.2/ojs/index.php/reinfo/article/view/171>>. Acesso em: 28 dez. 2014.

ARAÚJO, E. E. Revocação (recall) e precisão (precision) no SDI/CIN/CNEN. **Ciência da Informação**, v. 8, n. 1, 1979. Disponível em: <<http://revista.ibict.br/ciinf/index.php/ciinf/article/viewArticle/1545>>. Acesso em: 29 set. 2014.

ARAÚJO, V. M. R. H.; FREIRE, I. Conhecimento para o desenvolvimento: reflexões para o profissional da informação. **Informação & Sociedade: Estudos**, v. 9, n. 1, 1999. Disponível em: <<http://www.isafreire.pro.br/ARAUJOeFREIREConhe.pdf>>. Acesso em: 24 nov. 2014.

BARRETO, A. A condição da informação. **São Paulo em Perspectiva**, v. 16, n. 3, p. 67-74, 2002. Disponível em: <http://www.scielo.br/scielo.php?pid=S0102-88392002000300010&script=sci_arttext>. Acesso em: 15 fev. 2014.

_____. Mudança estrutural no fluxo do conhecimento: a comunicação eletrônica. **Ciência da Informação**, v. 27, n. 2, p. 122-127, 1998. Disponível em: <<http://www.scielo.br/pdf/ci/v27n2/barreto.pdf>>. Acesso em: 27 jun. 2014.

BENTO, A. Como fazer uma revisão da literatura: Considerações teóricas e práticas. **Revista JA** (Associação Acadêmica da Universidade da Madeira), n. 65, p. 42-44, 2012. Disponível em: <<http://www3.uma.pt/bento/Repositorio/Revisaodaliteratura.pdf>>. Acesso em: 10 out. 2014.

BERRY, M. J. A.; LINOFF, G. S. **Data mining techniques: for marketing, sales, and customer relationship management**. John Wiley & Sons, 2004.

BLATTMANN, U.; SILVA, F. C. C. Colaboração e interação na web 2.0 e biblioteca 2.0. **Revista ACB: Biblioteconomia em Santa Catarina**, v. 12, n. 2, 2007. Disponível em: <<http://revista.acbsc.org.br/racb/article/view/530>>. Acesso em: 29 nov. 2014.

BRANDT, M. B.; BRASCHER, M. B. Folksonomia: esquema de representação do conhecimento? **TransInformação**, Campinas, v. 22, n. 2, p. 111-121, 2010. Disponível em: <<http://repositorio.unb.br/handle/10482/12891>>. Acesso em: 11 out. 2014.

CAMPOS, L. F. B. Web 2.0, Biblioteca 2.0 e ciência da informação: um protótipo para disseminação seletiva de informação na Web utilizando mashups e feeds RSS. IN: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 8, Bahia, *Anais...* 2007. Disponível em: <<http://www.enancib.ppgci.ufba.br/artigos/GT2--232.pdf>>. Acesso em: 30 nov. 2014.

CAMPOS, M. L. A.; GOMES, H. E. Taxonomia e Classificação: o princípio de categorização. **DataGramZero**, v. 9, n. 4, 2008. Disponível em: <http://www.dgz.org.br/ago08/Art_01.htm>. Acesso em: 13 abr. 2014.

CARRILHO JUNIOR, J. R. **Desenvolvimento de uma metodologia para mineração de textos**. Orientador: Emmanuel Piseces Lopes Passos. – 2007. 96f.; 30 cm. Dissertação (Mestrado em Engenharia Elétrica) – Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, 2007. Disponível em: <http://www.maxwell.vrac.puc-rio.br/Busca_etds.php?strSecao=resultado&nrSeq=11675@1&msg=28#>. Acesso em: 15 dez. 2014.

CATARINO, M. E.; SOUZA, T. B. A representação descritiva no contexto da web semântica. **Transinformação**, Campinas, v. 24, n. 2, p. 77-90, 2012. Disponível em: <http://www.scielo.br/scielo.php?pid=S0103-37862012000200001&script=sci_arttext>. Acesso em: 13 nov. 2014.

_____, BAPTISTA, A. A. Folksonomia: um novo conceito para a organização dos recursos digitais na Web. **DataGramZero**, v. 8, n. 3, 2007. Disponível em: <http://www.dgz.org.br/jun07/Art_04.htm>. Acesso em: 25 nov. 2014.

CENDÓN, B. V. [et al.]. **Ciência da informação e biblioteconomia: novos conteúdos e espaços de atuação**. Belo Horizonte: Editora UFMG, 2005.

DA SILVA, M. B. **A teoria da classificação facetada na modelagem de dados em bancos de dados computacionais**. 2011. 168f. Dissertação (Mestrado) – Centro de Ciências Sociais Aplicadas, Universidade Federal da Paraíba, João Pessoa, 2011. Disponível em: <<http://rei.biblioteca.ufpb.br/jspui/bitstream/123456789/131/1/MBS01022013.pdf>>. Acesso em: 12 fev. 2014.

_____; DIAS, G. A. O Sistema de Automação em Bibliotecas *OpenBIBLIO* aplicado à disciplina Automação em Bibliotecas. **Biblionline** (João Pessoa), v. 6, p. 53-71, 2010. Disponível em: <<http://periodicos.ufpb.br/ojs2/index.php/biblio/article/view/4903>>. Acesso em: 11 out. 2014.

_____; NEVES, D. A. B. Estudo sobre o uso da teoria da classificação facetada em banco de dados. In: ENCONTRO NACIONAL DE PESQUISA EM PÓS-GRADUAÇÃO EM CIÊNCIA DA INFORMAÇÃO, 11. 2010, *Anais...* Florianópolis: UFSC, 2010. p.1 - 20. Disponível em: <<http://repositorios.questoesemrede.uff.br/repositorios/handle/123456789/870>>. Acesso em: 21 nov. 2014.

DAVENPORT, T., PRUSAK, L. **Conhecimento empresarial**. Rio de Janeiro: Campus, 1999. 237p.

DAVID K.; GLEDSON S. **Folksonomia: características, funcionamento e aplicações**. Terra Fórum, 2005. Disponível em: <<http://biblioteca.terraforum.com.br/Paginas/folksonomia-caracteristicas-funcionamento-e-aplicacoes.aspx>>. Acesso em: 02 nov. 2013.

DIAS, E. W. Contexto digital e tratamento da informação. **DataGramaZero**, v. 2, n. 5, 2001. Disponível em: <http://www.dgz.org.br/out01/Art_01.htm>. Acesso em: 23 nov. 2014.

FERNÁNDEZ-MOLINA, J. Enfoques objetivo y subjetivo del concepto de información. **Revista Española de Documentación Científica**, v. 17, n. 3, p. 320-331, 1994. Disponível em: <<http://dialnet.unirioja.es/servlet/articulo?codigo=647777>>. Acesso em: 30 nov. 2014.

FIGUEIRA, R. **Mineração de dados e bancos de dados orientados a objetos**. 1998. 96f. Dissertação (Mestrado em Ciências da Computação) – Departamento de Ciência da Computação, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 1998.

FRANCELIN, M. M.; PELLEGGATTI, C. Filosofia da informação: reflexos e reflexões. **Transinformação**, v. 16, n. 2, 2012. Disponível em: <<http://www.scielo.br/pdf/tinf/v16n2/02.pdf>>. Acesso em: 15 mar. 2014.

FURGERI, S. **Representação de informação e conhecimento: estudo das diferentes abordagens entre a ciência da informação e a ciência da computação**. Campinas: PUC-Campinas, 2006. Disponível em <http://www.bibliotecadigital.puc-campinas.edu.br/tde_busca/arquivo.php?codArquivo=174>. Acesso em: 04 mar. 2014.

GIARDELLI, G. Os dados nunca mentem. Bem-vindos à era do big data. **Exame**. 2013. Disponível em: <<http://exame.abril.com.br/rede-de-blogs/pessoas-do-seculo-21/2013/05/03/os-dados-nunca-mentem-bem-vindos-a-era-do-big-data/>>. Acesso em: 12 jan. 2015.

GNU OPERATING SYSTEM. **Whats is GNU?** 2014. Disponível em: <<https://www.gnu.org/>>. Acesso em: 25 nov. 2014.

GOOGLE. **Google Discovery**. 2014. Disponível em: <<http://googlediscovery.com/>>. Acesso em: 25 nov. 2014.

_____. **Google imagens**. 2014. Disponível em: <<https://www.google.com/imghp?hl=pt-BR>>. Acesso em: 25 nov. 2014.

GONÇALVES, L. P. F. **Avaliação de ferramentas de mineração de dados como fonte de dados relevantes para a tomada de decisão: aplicação na rede União de supermercados, São Leopoldo-RS**. 2001. Tese de Doutorado. Universidade Federal Do Rio Grande Do Sul. Disponível em: <<http://www.lume.ufrgs.br/handle/10183/1994>>. Acesso em: 21 jan. 2015.

GOUVÊA, C.; LOH, S. Folksonomias: identificação de padrões na seleção de tags para descrever conteúdos. **Revista Eletrônica de Sistemas de Informação**, v. 6, n. 2, 2007. Disponível em: <<http://189.16.45.2/ojs/index.php/reinfo/article/view/214>>. Acesso em: 11 out. 2014.

GUEDES, R.; AFONSO, D.; MAGALHÃES, L. H. Mineração de opiniões de usuários na busca de conhecimento. **Revista das Faculdades Integradas Vianna Júnior**, v. 1, ed. especial, p. 99, 2010. Disponível em: <http://www.viannajunior.edu.br/files/uploads/20131001_141137.pdf>. Acesso em: 07 abr. 2015.

GUEDES, R. M.; DIAS, E. J. W. Indexação Social: Abordagem Conceitual. **Revista ACB: Biblioteconomia em Santa Catarina**, v. 15, n. 1, p. 39-53, 2010. Disponível em: <http://www.brapci.inf.br/repositorio/2010/06/pdf_fcb17df2cd_0010808.pdf>. Acesso em: 27 nov. 2014.

GUY, M.; TONKIN, E. Folksonomies: tidying up tags? **D-Lib Magazine**, v. 12, n. 1, Jan. 2006. Disponível em: <<http://www.dlib.org/dlib/january06/guy/01guy.html>> Acesso em: 13 fev. 2015.

JAVA. **O que é o Java?** 2014. Disponível em: <https://www.java.com/pt_BR/>. Acesso em: 25 nov. 2014.

JESUS, A. P.; MOSER, E. M.; OGLIARI, P. J. Data Mining aplicado na identificação do perfil dos usuários de bibliotecas para a personalização de sistemas web de recuperação e disseminação de informações. In: WORKSHOP DE CIÊNCIA DA COMPUTAÇÃO E SISTEMAS DA INFORMAÇÃO DA REGIÃO SUL, 1, 2004, Palhoça. *Anais...* Disponível em: <<http://www.lbd.dcc.ufmg.br/colecoes/til/2004/0016.pdf>>. Acesso em: 10 jan. 2015.

LANCASTER, F. W. **Indexação e resumos**: teoria e prática. 2. ed. Brasília: Briquet de Lemos, 2004.

LAKATOS, E. M.; MARCONI, M. A. **Metodologia do trabalho científico**. 7 ed. São Paulo: Atlas, 2007.

LARA, M. L. G. Linguagem documentária e terminologia. **Transinformação**, Campinas, v. 16, n. 3, p. 231-240, 2004. Disponível em: <<http://www.brapci.ufpr.br/documento.php?dd0=0000000394&dd1=c13f1>>. Acesso em: 31 fev. 2014.

LE COADIC, Y.-F. **A Ciência da Informação**. Brasília: Briquet de Lemos Livros, 1996.

LIMA, G. A. B. Interfaces entre a ciência da informação e a ciência cognitiva. **Ciência da Informação**, v. 32, n. 1, 2003. Disponível em: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0100-19652003000100008>. Acesso em: 18 ago. 2014.

MACHADO, L. S., BECKER, K., O Uso da Mineração de Dados na Web Aplicado a um Ambiente de Ensino a Distância. In: I WORKSHOP DE TESES E DISSERTAÇÕES EM BANCO DE DADOS. XIX SIMPÓSIO BRASILEIRO DE BANCO DE DADOS. *Anais...* Gramado. 2002. Disponível em: <<http://www.lbd.dcc.ufmg.br/colecoes/wtdbd/2002/012.pdf>>. Acesso em: 01 dez. 2014.

MAGALHÃES, L. H. **Uma análise de ferramentas para mineração de conteúdo de páginas Web**. Rio de Janeiro, 2008. 76 p. (COPPE/ UFRJ, Mestrado em Engenharia Civil, 2008) Dissertação - Universidade Federal do Rio de Janeiro, COPPE. Disponível em: <http://www.coc.ufrj.br/teses/mestrado/Novas_2008/teses/MAGALHAES_LH_08_t_M_int.pdf>. Acesso em: 15 dez. 2014.

MAIA, N. **Educadores de sucesso: o que é metodologia**. Disponível em: <<http://educadoresdesucesso.blogspot.com.br/2011/02/o-que-e-metodologia.html>>. Acesso em: 30 nov. 2014.

MAIMONE, G. D.; SILVEIRA, Naira C.; TÁLAMO, M. F. G. M. Reflexões acerca das relações entre representação temática e descritiva. **Informação & Sociedade: Estudos**, v. 21, n. 1, 2011. Disponível em: <<http://www.ies.ufpb.br/ojs2/index.php/ies/article/view/7367>>. Acesso em: 23 nov. 2014.

MANESS, J. Teoria da biblioteca 2.0: web 2.0 e suas implicações para as bibliotecas. **Informação & Sociedade: Estudos**, v. 17, n. 1, 2007. Disponível em: <<http://periodicos.ufpb.br/ojs/index.php/ies/article/view/831>>. Acesso em: 25 nov. 2014.

MARCONI, M. de A.; LAKATOS, E. M. **Fundamentos de metodologia científica**. 5. ed. São Paulo: Atlas, 2007.

MARTELETO, R. M. Informação, rede e redes sociais: fundamentos e transversalidades. **Informação & Informação**, v. 12, n. 1, 2007. Disponível em: <http://www.brapci.inf.br/repositorio/2010/07/pdf_691c714087_0011336.pdf>. Acesso em: 21 maio 2014.

MATSUURA, S. Produção de dados dobra a cada dois anos, diz consultoria do IDC. **O globo**, 2012. Disponível em: <<http://oglobo.globo.com/sociedade/tecnologia/producao-de-dados-dobra-cada-dois-anos-diz-consultoria-do-idc-5980214>>. Acesso 29 nov. 2014.

MEDEIROS, T. B.; DA SILVA, M. B.; SILVA, R. Z. L. Um estudo sobre a biblioteca das Faculdades FACENE/FAMENE: em foco, a seção de multimeios. **Biblionline** (João Pessoa), v. 10, p. 135-153, 2014. Disponível em: <<http://periodicos.ufpb.br/ojs/index.php/biblio/article/view/13472>>. Acesso em: 23 jan. 2015.

MEY, E. S. A. Da espiral do conhecimento à catalogação. **Revista de Biblioteconomia de Brasília**, v. 15, n. 2, p. 137-148, 1987. Disponível em: <<http://www.brapci.ufpr.br/documento.php?dd0=0000008297&dd1=3063b>>. Acesso em: 10 nov. 2014.

MOREIRA FILHO, J. L. ZAPPAROLI, Z. M. Adaptatividade para pesquisas em linguística de corpus. In: SÉTIMO WORKSHOP DE TECNOLOGIA ADAPTATIVA, 2013. *Anais...* EPUSP, 2013. Disponível em: <<http://lta.poli.usp.br/lta/publicacoes/artigos/2013/moreira-filho-e-zapparoli-2013-adaptatividade-para-pesquisas-em-linguistica-de-corpus/view>>. Acesso em: 30 nov. 2014.

MOURA, Maria Aparecida. Leitor-bibliotecário: interpretação, memória e as contradições da subjetividade. *Perspectivas em ciência da informação*, v. 9, n. 2, 2004. Disponível em: <<http://portaldeperiodicos.eci.ufmg.br/index.php/pci/article/view/357/0>>. Acesso em: 12 dez. 2014.

KURAMOTO, H. **Organização da informação: princípios e tendências**. Brasília: Briquet de Lemos, 2006.

NAVATHE, S. B.; ELMASRI, R. **Sistemas de banco de dados**. Sham Addison: São Paulo, 2005.

NICHOLSON, S.. O processo da bibliomineração: repositório de dados e mineração de dados para tomada de decisão em bibliotecas. *Transinformação*, v. 16, n. 3, 2012. Disponível em: <<http://periodicos.puc-campinas.edu.br/seer/index.php/transinfo/article/view/712>>. Acesso em: 12 dez. 2014.

NICOLAU, J. A.; PARANHOS, J. Notas sobre o conceito de inovação. *Textos de Economia*, v. 9, n. 1, p. 23-37, 2006. Disponível em: <<https://periodicos.ufsc.br/index.php/economia/article/viewFile/2302/1986>>. Acesso em: 27 set. 2014.

NOVO, R.; NEVES, J. M. S. Inovação na inteligência analítica por meio do Big Data: características de diferenciação da abordagem tradicional. In: WORKSHOP DE PÓS-GRADUAÇÃO E PESQUISA DO CENTRO PAULA SOUZA, 8, 2013, *Anais...* São Paulo, p. 32-44, 2013. Disponível em: <http://www.centropaulasouza.sp.gov.br/pos-graduacao/workshop-de-pos-graduacao-e-pesquisa/008-workshop-2013/trabalhos/desenvolvimento_de_tecnologia_e_sistemas/121191_32_44_FINAL.pdf>. Acesso em: 23 jan. 2015.

OLIVEIRA, G. B. Uma discussão sobre o conceito de desenvolvimento. *Revista da FAE*, v. 5, n. 2, p.37-48, 2002. Disponível em: <http://www.fae.edu/publicacoes/pdf/revista_da_fae/fae_v5_n2/uma_discussao_sobre.pdf>. Acesso em: 21 ago. 2014.

OLIVEIRA, V. P. Uma informação tácita. *DataGramZero*, v. 6, n. 3, 2005. Disponível em: <http://www.dgz.org.br/jun05/Art_04.htm>. Acesso em: 19 jul. 2014.

PERGAMUN. **Rede Pergamum**. 2014. Disponível em: <<http://www.pergamum.pucpr.br/redepergamum/>>. Acesso em 29 nov. 2014.

PIEDADE, M. **Introdução a teoria da classificação**. 2. ed. rev. e ampl. Rio de Janeiro: Interciência, 1983.

PINTO, V. B. Indexação documentária: uma forma de representação do conhecimento registrado. **Perspectivas em ciência da informação**, v. 6, n. 2, 2001. Disponível em: <<http://portaldeperiodicos.eci.ufmg.br/index.php/pci/article/view/423>>. Acesso em: 25 nov. 2014.

PRIMO, A. O aspecto relacional das interações na Web 2.0. **E-Compós (Brasília)**, v. 9, p. 1-21, 2007. Disponível em: <<http://www.compos.org.br/seer/index.php/e-compos/article/viewArticle/153>>. Acesso em: 23 nov. 2014.

PRODANOV, C. C.; FREITAS, E. C. **Metodologia do trabalho científico: métodos e técnicas da pesquisa e do trabalho acadêmico**. [recurso eletrônico] – 2. ed. – Novo Hamburgo: Feevale, 2013. Disponível em: <<https://www.feevale.br/cultura/editora-feevale/metodologia-do-trabalho-cientifico---2-edicao>>. Acesso em: 30 nov. 2014.

PYTHON BRASIL. **O que é Python?** 2014. Disponível em: <<http://wiki.python.org.br/>>. Acesso em: 25 nov. 2014.

RUBI, M. P. Os princípios da política de indexação na análise de assunto para catalogação. IN: FUJITA, M. S. L. (org.). **A indexação de livros: a percepção de catalogadores e usuários de bibliotecas universitárias**. São Paulo: Editora UNESP, 2009.

RUFINO, F. M.; JESUS, J. R.; DA SILVA, M. B. A web 2.0 em bibliotecas: vertentes de estudo. In: 3º SEMINÁRIO DE INFORMAÇÃO EM ARTE, *Anais...* 2013. Rio de Janeiro: Fundação Casa de Rui Barbosa, 2013. Disponível em: <<http://redarterj.com/wp-content/uploads/2014/11/A-WEB-2.0-em-Bibliotecas.pdf>>. Acesso em: 12 jan. 2015.

SANTOS, P. L. V. C.; SANT'ANA, R. C. G. Transferência da informação: análise para valoração de unidade de conhecimento. **DataGramZero**, v. 3, n. 2, 2002. Disponível em: <http://www.dgz.org.br/abr02/Art_02.htm>. Acesso em: 22 jan. 2014.

SARACEVIC, T. Tecnologia da informação, sistemas de informação e informação como utilidade pública. **Ciência da informação**, v. 3, n. 1, 1974. Disponível em: <<http://revista.ibict.br/ciinf/index.php/ciinf/article/view/168>>. Acesso em: 30 jun. 2014.

SETZER, V. W. Dado, informação, conhecimento e competência. **DataGramZero**, n. 0, 1999. Disponível em: <http://www.dgz.org.br/dez99/Art_01.htm>. Acesso em: 05 jun. 2014.

SFERRA, H. H.; CORRÊA, A. M. C. J. Conceitos e aplicações de data mining. **Revista de ciência & tecnologia**, v. 11, n. 22, 2003. Disponível em: <<http://www.unimep.br/phpg/editora/revistaspdf/rct22art02.pdf>>. Acesso em: 10 jan. 2015.

SILVA, E. L.; MULLER, M. S. A presença do jornalismo investigativo no “Jornal do Tocantins”. IN: INTERCOM – SOCIEDADE BRASILEIRA DE ESTUDOS INTERDISCIPLINARES DA COMUNICAÇÃO, 34, 2011. *Anais...* Recife, PE, 2011. Disponível em: <<http://www.intercom.org.br/papers/nacionais/2011/resumos/r6-2320-1.pdf>>. Acesso em: 30 nov. 2014.

SILVA, E. L.; MENEZES, E. M. **Metodologia da pesquisa e elaboração de dissertação**. 4. ed. rev. atual. – Florianópolis: UFSC, 2005. 138p. Disponível em: <[https://projetos.inf.ufsc.br/arquivos/Metodologia de pesquisa e elaboracao de teses e dissertacoes_4ed.pdf](https://projetos.inf.ufsc.br/arquivos/Metodologia_de_pesquisa_e_elaboracao_de_teses_e_dissertacoes_4ed.pdf)>. Acesso em 30 nov. 2014.

SILVA, L. V. S. **Mineração de dados em Sistema de Recomendação Web**. 2011. Monografia. Faculdade de tecnologia de São José dos Campos. Disponível em: <http://fatecsjc.edu.br/trabalhos-de-graduacao/wp-content/uploads/2012/04/TG_INTRODU%C3%87%C3%83O_MINERA%C3%87%C3%83O_DE_DADOS_EM_REDES-SOCIAIS_LUCAS_VINICIUS.pdf>. Acesso em: 12 dez. 2014.

SILVA, M. A. C.; SANTOS, J. L. C. Análise da gestão do conhecimento no INPA baseada em práticas gerenciais. **DataGramZero**, v. 9, n.4, ago. 2008. Disponível em: <http://www.dgz.org.br/ago08/Art_03.htm>. Acesso em: 12 nov. 2014.

SILVA, M. P. S. Mineração de Dados: Conceitos, Aplicações e Experimentos com Weka. **Livro da Escola Regional de Informática**. Porto Alegre: Sociedade Brasileira de Computação, 2004, v. 1, p. 1-20. Disponível em: <<http://www.lbd.dcc.ufmg.br/colecoes/erirjes/2004/004.pdf>>. Acesso em: 14 jul. 2014.

SILVA, M. R.; FUJITA, M. S. L. A prática de indexação: análise da evolução de tendências teóricas e metodológicas. **Transinformação**, v. 16, n. 2, 2012. Disponível em: <<http://www.scielo.br/pdf/tinf/v16n2/03.pdf>>. Acesso em: 13 out. 2014.

SOMASUNDARAM, G; SHRIVASTAVA, A. **Armazenamento e Gerenciamento de Informações: Como armazenar, gerenciar e proteger informações digitais**. Bookman, 2011.

SOUZA, R. R. Sistemas de recuperação de informações e mecanismos de busca na web: panorama atual e tendências. **Perspectivas em Ciência da Informação** (Impresso), Belo Horizonte, v. 11, n. 2, p. 161-173, 2006. Disponível em: <<http://www.scielo.br/pdf/pci/v11n2/v11n2a02>>. Acesso em: 24 set. 2014.

TOMAÉL, M. I.; ALCARA, A. R.; DI CHIARA, I. G. Das redes sociais à inovação. **Ciência da Informação**, Brasília, DF, v. 34, n. 2, p. 93-104, 2005. Disponível em: <<http://revista.ibict.br/ciinf/index.php/ciinf/article/view/642>> Acesso em: 08 ago. 2014.

UNISIST. Princípios de indexação. **Revista da Escola de Biblioteconomia da UFMG**, Belo Horizonte, v.10, n.1, p.83-94, 1981. Disponível em: <www.brapci.ufpr.br/download.php?dd0=15948>. Acesso em: 12 dez. 2014.

UOL SAC. **Incidentes de segurança e uso abusivo da rede**. 2015. Disponível em: <<https://sac.uol.com.br/info/cartilha/incidentes/sec2.jhtm>>. Acesso em 02 mar. 2015.

VALE, M. C. Conceito de Desenvolvimento. **Acta pediátrica**, v. 36, n. 2/3, p. 105-108, 2005. Disponível em: <<http://actapediatrica.spp.pt/article/viewFile/4858/3657>>. Acesso em: 24 set. 2014.

VITAL, L. P.; CAFÉ, L. M. A. Ontologia e taxonomia: diferenças. **Perspectivas em Ciência da Informação**, v. 16, n. 2, p. 115-130, 2011. Disponível em: <<http://portaldeperiodicos.eci.ufmg.br/index.php/pci/article/view/200>>. Acesso em: 25 nov. 2014.

WAL, T. V. **Folksonomy definition and Wikipedia**. 2006. Disponível em: <<http://www.vanderwal.net/random/entrysel.php?blog=1750>>. Acesso em: 22 nov. 2014.

WURMAN, R. S. **Ansiedade de informação**. São Paulo: Cultura Editores Associados, 1991. 239 p.

YONG, C. S. **Banco de Dados – Organização, sistemas e administração**. São Paulo: Editora Atlas S. A, 1986.