

2 Teoria da Classificação

O ato de classificar não é uma tarefa simples, por mais que inventemos regras para classificação de um dado, essa classificação pode ser subjetiva devido à intervenção humana no ato da criação das categorias, bem como na busca deste dado. Desta maneira, dificilmente conseguiremos classificar um dado de forma única, permitindo que a busca seja coerente a todos os seus utilizadores.

Estudiosos viram a necessidade de construir sistemas únicos de classificação que abrangessem todas as áreas do conhecimento, tentando possibilitar uma organização geral para os dados. Neste capítulo veremos algumas das teorias de classificação desenvolvidas ao longo dos séculos e uma breve análise positiva e negativa de cada abordagem.

2.1. Classificação Decimal de Dewey (CDD)

Em 1876, o bibliotecário norte-americano Melvil Dewey publicou o sistema de Classificação Decimal, que tem sido modificado e expandido desde sua criação. Até o ano de 2007 ocorreram 22 grandes revisões. Hoje, o CDD é o sistema de classificação bibliográfica mais utilizada no mundo.

Esse sistema de classificação é dividido em dez classes principais, que juntas tentam cobrir todas as áreas do conhecimento existentes. Cada classe principal (ver tabela 1) é dividida em outras dez divisões; cada divisão, em mais dez seções.

A classe 000 é a mais geral e não é utilizada para um assunto específico. As demais classes englobam uma determinada área do conhecimento e são organizadas por disciplinas e não por assuntos, pois um mesmo assunto pode estar contido em mais de uma classe.

Classes Principais de Dewey

000	Computers, information & general reference
100	Philosophy & psychology
200	Religion
300	Social sciences
400	Language
500	Science
600	Technology
700	Arts & recreation
800	Literature
900	History & geography

Tabela 1 As dez classes principais da classificação decimal de Dewey

Embora pareça limitado, nem todos os números para divisões e seções foram utilizados. O primeiro número representa a classe principal do domínio; o segundo, a divisão da classe principal; e o terceiro, as seções.

A CDD em sua notação foi convencionada que não possuiria menos de três dígitos em sua representação e que um ponto (.) seria postado antes do quarto dígito, caso este ocorra, para tentar tornar a leitura mais clara. Após o ponto decimal, a divisão por dez continua até atingir a precisão adequada, vejamos um exemplo da CDD na figura 1.

600	Technology (Applied Sciences)
630	Agriculture and related technologies
636	Animal husbandry
636.7	Dogs
636.8	Cats

Figura 1 Exemplo da Classificação de Dewey.

Neste exemplo, podemos notar que a CDD é estruturada hierarquicamente, de forma que todos os tópicos (a partir das dez categorias principais) estejam ligados a outros mais gerais. Um exemplo real deste sistema é o Dewey Browse¹, que foi criado por Gail Shea Grainger e demonstra como seria a utilização deste método na organização de dados na internet.

¹ <http://www.deweybrowse.org/>

A grande vantagem desse sistema de classificação é a sua simplicidade. Uma notação simples é fundamental para a sua difusão. Outra vantagem é a sua infinita expansão vertical, ou seja, quando novas subdivisões são introduzidas não há mudança na alocação dos números de sua classificação. É importante afirmar que uma subdivisão é feita a partir de uma divisão já existente, e não pela introdução de uma nova divisão ao sistema.

A desvantagem desse sistema está em sua difícil adaptabilidade, que em níveis mais altos de abstração possui categorias já definidas. Outra desvantagem, proveniente desde os primórdios de sua criação por Dewey, são as categorias que englobam o conhecimento essencialmente de sua experiência [1], enquanto os outros métodos de classificação utilizam especialistas para descrever uma categoria.

2.2. Classificação Decimal Universal (CDU)

A Classificação Decimal Universal foi adaptada por Paul Otlet e Henri La Fontaine da Classificação Decimal de Dewey (CDD) e foi publicada pela primeira vez na França entre os anos de 1904 e 1907. Assim como a CDD, a CDU desde sua criação tem sido desenvolvida e revisada se tornando um sistema flexível e efetivo para organizar arquivos bibliográficos para todos os tipos de informações e mídias, ultrapassando a CDD em tamanho e força; contudo a estrutura lógica permanece a mesma.

A diferença entre a CDD e a CDU está no modo como os assuntos de determinada obra podem ser correlacionados. A CDU possui uma tabela de símbolos, como o sinal de mais, ponto e vírgula, travessão, entre outros que podem ser divididos em dois grupos: dependentes (ponto de vista; características gerais de pessoas e materiais) e independentes (auxiliares comuns de língua; forma; lugar; raças e povos; tempo).

Por ser semelhante à CDD, a CDU possui praticamente os mesmos prós e contras em seu sistema. Entretanto, a CDU possui uma independência perante todas as expressões idiomáticas já que a CDD foi baseada na cultura americana, o que facilita a pesquisa e a troca de informação em âmbito mundial.

2.3. Classificação da Biblioteca do Congresso (LCC)

Esta classificação é a mais utilizada nas bibliotecas acadêmicas e de pesquisas nos Estados Unidos. Desenvolvida por Herbert Putnam com os conselhos de Charles Ammi Cutter em 1987, esta classificação não descreve as áreas do conhecimento como um todo, mas assuntos de livros em uma biblioteca. Entretanto, suas categorias são divididas abstratamente.

Uma importante característica da LCC é a sua notação, definida como “uma série de símbolos que representam os nomes de uma classe ou a qualquer divisão ou subdivisão de uma classe que forma uma lógica para a classificação” [2]. A notação é mista, composta por letras maiúsculas e numerais arábicos.

O sistema divide todo o conhecimento em 21 classes principais e cada uma delas é identificada por uma letra do alfabeto, ver tabela 2.

Classes da Biblioteca do Congresso

A	Generalidades
B	Filosofia. Psicologia. Religião
C	Ciências auxiliares da História
D	História Geral e História da Europa
E	História da América (Generalidades e EUA)
F	História da América (outros países)
G	Geografia. Antropologia. Lazer
H	Ciências Sociais
J	Ciências Políticas
K	Direito
L	Educação
M	Música e Musicologia
N	Belas Artes
P	Linguística e Literatura
Q	Ciências
R	Medicina
S	Agricultura
T	Tecnologia
U	Ciência Militar

V	Ciência Naval
Z	Bibliografia, Ciências da Informação e das Bibliotecas

Tabela 2 Classes da Biblioteca do Congresso (LLC).

A maioria dessas classes é subdividida em subclasses identificadas por duas, ou, às vezes, três letras (ver exemplo da LCC [1] na figura 3). Tópicos individuais são quebrados em lugares, períodos, periódicos, biografias etc, e a cada um desses tópicos, um número entre um e quatro dígitos é atribuído, note que as letras I, O, W, X e Y não são utilizadas.

T – Technology
TJ – Mechanical Engineering
248 – Mechanical Models
249 – Erecting work
250 – Prime movers in general

Figura 2 Exemplo da LLC.

Ao contrário da CDD a classificação da LLC não é hierárquica, ou seja, um conceito não é desenvolvido a partir do mais geral para o mais específico, a sua notação apenas segue uma seqüência convencional [1] entre tópicos de um determinado assunto.

A vantagem deste método diz respeito às categorias da LCC, criadas por especialistas nas áreas do conhecimento de acordo com a necessidade da biblioteca do congresso.

A desvantagem deste sistema está no uso de uma linguagem controlada que, por ser arbitrária, pode parecer não natural. Além disso, foi originalmente criada para atender a apenas uma biblioteca, a LCC.

2.4. Classificação Facetada (Colon Classification)

A teoria da classificação facetada foi desenvolvida pelo matemático e bibliotecário indiano Shiyali Ramamrita Ranganathan em 1933. Insatisfeito com os sistemas de classificação Decimal de Dewey e a Classificação Decimal Universal, Ranganathan decidiu criar seu próprio sistema de classificação,

visando melhorar a representação de assuntos complexos e/ou compostos. Além deste sistema de classificação, Ranganathan produziu outras obras importantes para a teoria da classificação, como por exemplo: *Elements of Library Classifications* (1945; 1962) e *Prolegomena to Library Classification* (1967).

A classificação de dois pontos também é conhecida como analítico-sintética, já que decompos os assuntos/temas em seus conceitos básicos (análise) e os combinamos para descrever um assunto/tema (síntese). Assim como as classificações já citadas, ela possui classes de assuntos pré-determinadas que podem ser usadas em combinação entre elas e por meio de marcadores como, por exemplo, o *colon*.

Ranganathan identificou cinco classes principais, também conhecidas como categorias fundamentais, representadas pelo mnemônico PMEST, cujas facetas se enquadram. A categoria *Personality* é considerada a faceta principal, característica que distingue o assunto, a categoria *Matter* destina-se a identificação do material físico do qual um objeto pode ser composto, a categoria *Energy* é responsável pela identificação da atividade ou processo que ocorra com o assunto, a categoria *Space* identifica aonde o assunto acontece ou existe, localização geográfica e a categoria *Time* que identifica quando o assunto ocorre, período de tempo associado ao assunto.

“A classificação facetada compreende princípios e técnicas para os sistemas de organização e recuperação da informação. Um sistema facetado reconhece muitos aspectos em um único assunto, e tenta sintetizá-los de maneira mais adequada. Ranganathan mostrou que a relação entre assuntos, feita por meio do símbolo de dois pontos, pode ser infinita, provando que o conhecimento pode ser multidimensional e que essas relações podem tomar rumos diferentes, dependendo da síntese entre vários conceitos múltiplos.” [3]

A elaboração de facetas é feita a partir da inspeção do assunto com o propósito de identificar conceitos e termos que possam ser particionados em facetas. Além disso, cada faceta pode ser subdividida em subfacetas. A seguir, vamos exemplificar como a esta classificação funciona. Dada a frase "*Circulation of periodicals in University Libraries in India up to the 1970s*", a classificação resultante é: 234;46:6.44'N7, para melhor compreensão vamos decompor este resultado e mostrar como foi aplicado a classificação de dois pontos.

O número 2 indica a classe básica “*Library Science*”, o número 3 indica que é qualquer biblioteca acadêmica e o número 4 indica que é uma biblioteca universitária (*Personality*).

;46 = O ponto e vírgula indica o valor da *matter*. Neste caso, 46 corresponde a periódicos.

:6 = O símbolo de dois pontos indica um valor de *Energy*. No domínio da *Library Science*, facetas que estão relacionadas ao conceito de *Energy*, descrevem ações, tais como, catalogar (55), circulação (6) entre outras.

.44 = O ponto indica um valor de *Space*. Aonde o número 44 é atribuído a Índia.

‘N7 = O apóstrofo indica um valor de tempo. A letra inicial indica o século (N=1900-1999), enquanto o número indica a década.

A grande vantagem da classificação facetada são as diversas possibilidades de navegação e busca da mesma informação, oferecendo ao usuário uma variedade de estratégias de busca e modelos conceituais. Ao mesmo tempo, ela também permite identificar cada um dos sentidos a que se refere um assunto. Ao contrário das demais classificações supracitadas, que são enumerativas, a classificação facetada é analítico-sintética, ou seja, possui uma estrutura multidimensional e de fácil adaptabilidade.

2.5. Conclusão

Esta seção serve de apoio na identificação de fatores que influenciam na classificação de dados. Torna-se evidente a necessidade de uma taxonomia ser adaptável e ter facilidade de expansão como foi citado na CDD. Possuir internacionalização para poder ser inteligível a nível mundial e não regional como foi visto na CDU. Possuir um vocabulário não controlado como o da LCC para que não haja a necessidade de existir especialistas em diversas áreas do conhecimento para manter essa organização.

Além disso, há a necessidade de criar taxonomias que possam agrupar assuntos complexos (multidimensionais) como os citados por Ranganathan quando descreve os fatores PMEST.