

ESTUDO DAS SIMILARIDADES ENTRE A TEORIA MATEMÁTICA DA COMUNICAÇÃO E O CICLO DOCUMENTÁRIO

*William Guedes**
*Rogério Henrique de Araújo Júnior***

RESUMO

Trata do estudo, no âmbito da ciência da informação, da teoria matemática da comunicação proposta por Claude Shannon em 1948 e as suas similaridades com o ciclo documentário no que tange às etapas que vão desde a seleção, armazenamento e tratamento técnico de documentos, até a recuperação da informação pelos usuários. O tema é introduzido apresentando quatro vertentes dentre as inúmeras abordagens dos sistemas de comunicação na ciência da informação, representadas pela: 1. comunicação científica; 2. dicotomia entre linguagens natural e documentária; 3. inserção da teoria matemática da comunicação na ciência da informação e 4. ciclo documentário, que tem na recuperação da informação a etapa mais significativa sob o ponto de vista de um sistema de comunicação. Apresenta a definição, o escopo e as características da teoria matemática da comunicação, reforçando o seu caráter pioneiro ao propor, sob bases sólidas, o cálculo da quantidade de informação que pode transitar por um canal de comunicação. Aborda também questões relacionadas com neutralidade do canal de comunicação que estão presentes em discussões na ciência da informação, tomando como exemplo a valoração do documento segundo a teoria arquivística. Finalmente, são delimitadas as similaridades da teoria matemática da comunicação com o ciclo documentário, onde um documento pode ser visto também como um canal de comunicação, assim como qualquer suporte que apresente informação a ser tratada sob o ponto de vista documentário.

Palavras-chave: Ciência da informação. Teoria matemática da comunicação. Ciclo documentário. Sistemas de comunicação.

* Mestre em Ciências da Informação pela Universidade de Brasília, Brasil. Analista de Finanças e Controle da Controladoria-Geral da União, Brasil. E-mail: wguedes@outlook.com.

** Doutor em Ciências da Informação pela Universidade de Brasília, Brasil. Professor da Faculdade de Ciência da Informação da Universidade de Brasília, Brasil. E-mail: araujojr@unb.br.

I INTRODUÇÃO

Os sistemas de comunicação têm sido estudados por pesquisadores de diferentes áreas de conhecimento, a exemplo de ciência da informação e engenharia. Na ciência da informação destacamos quatro vertentes de abordagem para os sistemas de comunicação.

A primeira aponta para a comunicação de pesquisas desenvolvidas, onde o seu estudo é parte integrante do desenvolvimento da

ciência. Segundo Perucchi e Araújo Junior (2011) a legitimação do conhecimento se dá pela divulgação, verificação e comprovação da produção de conhecimento, processo que só é possível por meio dos sistemas de comunicação científica, que por sua vez encerram indicadores importantes para a medida da atividade de pesquisa, ou seja, a contabilização do número de artigos e trabalhos científicos publicados e o impacto dessas publicações entre os pares.

A segunda vertente pode ser cotejada na diferença entre linguagem natural e

documentária. A linguagem natural pode ser conceituada como a própria comunicação humana com a utilização da palavra como forma de expressão e comunicação entre indivíduos (emissor e receptor). De outra maneira, é a forma de interação humana realizada através do uso de signos, dada por meio da expressão escrita ou oral. Já a linguagem documentária, no âmbito da ciência da informação, é aquela que objetiva traduzir semanticamente o conteúdo dos documentos em uma linguagem mais especializada, criada, exclusivamente, para apoiar o processo de busca e recuperação da informação, por meio da representação do conteúdo dos documentos.

A terceira vertente de abordagem trata da adaptação da teoria matemática da comunicação, que representa uma das inúmeras contribuições recolhidas pela Ciência da Informação para delimitar o seu principal objeto de pesquisa, o fenômeno da informação. A teoria matemática da comunicação se encaixa, segundo texto clássico de Wersig e Neveling (1975), na abordagem da mensagem – onde a informação é frequentemente usada como sinônimo de mensagem, pois o seu conteúdo é medido pela quantidade de informação que pode percorrer um canal de comunicação. Desse modo, a efetividade da comunicação vai ser dada pela relação entre o resultado esperado pelo emissor e o que ocorre no receptor.

A quarta vertente de abordagem dos sistemas de comunicação é representada pelo ciclo documentário que está configurado como um processo composto de operações que vão desde a seleção, tratamento técnico (representação do conteúdo dos documentos apoiada pela utilização das linguagens documentárias), armazenamento em sistema de informação, passando pela criação de produtos do processamento dos documentos, até a recuperação da informação pelos usuários, que é a etapa mais significativa sob o ponto de vista de um sistema de comunicação.

Assim sendo, o objetivo deste artigo é mostrar que existem similaridades claras entre as abordagens das vertentes da teoria matemática da comunicação e do ciclo documentário. Em ambos os casos a problemática que associa a escolha, organização, codificação e transmissão da mensagem, questões centrais da teoria matemática da comunicação, também estão

presentes na consideração das etapas do ciclo documentário, notadamente no que se refere ao armazenamento, indexação e recuperação da informação.

2 TEORIA MATEMÁTICA DA COMUNICAÇÃO

Publicado em 1948, um trabalho do engenheiro e matemático norte-americano Claude Shannon teve especial importância. Pela primeira vez alguém conseguiu calcular a quantidade de informação que poderia transitar por um canal de comunicação. Isso foi possível a partir análise estatística da frequência de ocorrência de letras nas palavras de um idioma. Shannon estudou a distribuição das letras na formação de mensagens escritas no idioma inglês. Segundo as regras ortográficas, a probabilidade de umas letras ocorrerem em sequência a outras não é igual em todos os casos. A letra E ocorre mais frequentemente do que o Q, e a sequência TH mais frequentemente do que XP, etc. Shannon, (1948, p. 5). Ele concluiu que aproximadamente metade do que se escreve em inglês é determinado pela estrutura da linguagem, sendo apenas a outra metade de livre escolha do redator. Ou seja, metade do que se escreve é redundância imposta pela estrutura do idioma (neste caso, o inglês). Na língua portuguesa, a redundância deve ser ainda maior, segundo Fidalgo (2004):

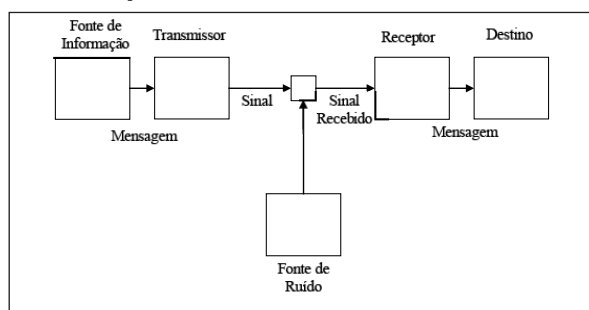
De facto, as partes redundantes da mensagem constituem algo que não traz novidade e, portanto, serão desnecessárias nesse sentido. A não necessidade da citada redundância do artigo definido em português em género e número torna-se clara quando comparado com o artigo definido inglês “the” que não conhece nem género nem número. Tendo isto em conta, e ainda, sobretudo, a personalização das formas verbais que existem na língua portuguesa em muito maior grau que na língua inglesa (às seis formas pessoais dos verbos portugueses – de *eu amo* a eles *amam* correspondem no geral apenas duas nos verbos ingleses – *I, you, we, they love, he loves*), é quase certo que a percentagem de redundância em português é superior aos 50% de redundância que Shannon e Weaver atribuem à língua inglesa, em que apenas

metade das letras ou das palavras que escrevemos ou dizemos são de livre escolha de quem fala, e que a outra metade é ditada pela estrutura estatística da língua. (FIDALGO, 2004, p.5, grifo do autor).

A redundância significa desperdício do canal de comunicação, porque o acréscimo do número de palavras em decorrência das regras do idioma não implica mais informação transmitida. Mas ela não deve ser totalmente eliminada, porque contribui para a redução do ruído na comunicação.

O sistema genérico de comunicação estudado por Shannon está representado na Figura 1.

Figura 1 - Diagrama de um sistema geral de comunicação



Fonte: Shannon (1948)

Ele contém cinco partes:

- Uma fonte de informação: produz a mensagem (ou sequência de mensagens) que será comunicada ao receptor. A mensagem pode ser de vários tipos: uma sequência de letras, como em um sistema telegráfico; um sinal de rádio ou de uma ligação telefônica; um sinal de televisão.
- Um transmissor: é o elemento que opera a mensagem de modo a produzir um sinal que seja adequado para ser transmitido pelo canal.
- O canal: é o meio usado para transmitir o sinal do transmissor ao receptor.
- O receptor: realiza a operação inversa à do transmissor, reconstruindo a mensagem a partir do sinal recebido.
- O destinatário: é a pessoa (ou máquina) à qual a mensagem se destina (SHANNON, 1948).

Essa abordagem, essencialmente técnica, não considera o significado das palavras, nem do texto que elas formam.

O problema fundamental da comunicação é o de reproduzir em um ponto, ou exatamente ou aproximadamente, uma mensagem selecionada em outro ponto. Frequentemente as mensagens têm significado, isto é, elas se referem a ou são correlacionadas de acordo com algum sistema com certas entidades físicas ou conceituais. Estes aspectos semânticos da comunicação são irrelevantes para o problema de engenharia. (SHANNON, 1948, p. 1).

A irrelevância dos aspectos semânticos da comunicação para a engenharia tem uma razão muito simples: um canal de comunicação deve ser neutro, permitindo o trânsito de qualquer mensagem, independentemente de seu significado.

Essa neutralidade do canal de comunicação está também presente em discussões na ciência da informação. Duranti (1993) faz uma retrospectiva histórica da valoração de um documento segundo a teoria arquivística, e relata como o conceito de avaliação foi acompanhado das ideias de imparcialidade, objetividade e profissionalismo na literatura da Alemanha, dos Estados Unidos, da Inglaterra e do Canadá ao longo do século XX. “A avaliação precisa estar baseada em conhecimento derivado de análise e deve ser destinada a fornecer um quadro completo da sociedade [...]” (Duranti, 1993, p. 330). O arquivo, portanto, é o canal de comunicação que permite aos registros transitarem entre gerações. Deve ser neutro, para não induzir, privilegiar ou distorcer os fatos, pois só assim a análise de boa qualidade será possível. “A profissão arquivística tem uma responsabilidade vital com as futuras gerações, de deixar que elas entendam e julguem nossa sociedade com base nos documentos que produziu [...]” (Duranti, 1993, p. 343). O arquivista não deve confundir seu papel de preservador dos registros com o de criador de valor (ao avaliar e selecionar documentos, descartando aqueles que julgasse sem utilidade, por exemplo, estaria alterando o conteúdo do acervo que deve preservar).

3 DIFERENTES ABORDAGENS PARA OS PROBLEMAS DE COMUNICAÇÃO

Em 1949, o matemático Weaver (1949), publicou o texto *Recent Contributions to The Mathematical Theory of Communication* no qual apresentou interpretações para o trabalho de Shannon e forneceu exemplos que facilitaram sua compreensão. Os dois artigos foram então publicados juntos em forma de livro, na obra assinada por ambos e intitulada *The Mathematical Theory of Communication*.

Em seu texto, Weaver dividiu os problemas de comunicação em três níveis:

Nível A – Problema técnico: quão precisa pode ser a transmissão de símbolos?

Nível B – Problema semântico: quão precisamente os símbolos transmitidos carregam o significado desejado?

Nível C – Problema da efetividade: quão efetivamente o significado recebido afeta a conduta da maneira desejada? (WEAVER, 1949, p. 1).

O problema técnico está ligado, por exemplo, à transmissão de mensagens escritas por correio eletrônico, de voz por telefone ou de imagens e sons por televisão. O problema semântico está relacionado a quão próximas são as interpretações de significado na transmissão e na recepção – o que implica comparar se o que o emissor quis transmitir foi compreendido pelo receptor da mensagem. A efetividade da comunicação é a relação entre o resultado esperado pelo emissor e o que ocorre no receptor, ou seja, é o cotejamento entre o impacto pretendido pelo emissor com sua mensagem e o que de fato ocorreu com o usuário da informação.

Weaver entende que o problema de nível A não está restrito às especificidades técnicas de engenharia, porque o conteúdo filosófico da comunicação, associado aos níveis B e C, não é transmitido quando o nível A é insatisfatório:

A teoria matemática dos aspectos de engenharia da comunicação, desenvolvida principalmente por Claude Shannon no Bell Telephone Laboratories, reconhecidamente se aplica em primeira instância apenas ao problema A, ou seja, o problema técnico de precisão da transferência de

vários tipos de sinais entre o emissor e o receptor. Mas a teoria tem, penso eu, um profundo significado que prova que o parágrafo anterior é seriamente impreciso. Parte da importância da nova teoria vem do fato de que os níveis B e C, acima, somente podem fazer uso da precisão daqueles sinais que venham a ser possíveis quando analisados ao nível A. Assim, quaisquer limitações descobertas na teoria no Nível A se aplicam necessariamente aos níveis B e C. Mas uma maior parte da importância vem do fato de que as análises no nível A revelam que este se sobrepõe aos outros níveis mais do que possivelmente se poderia ingenuamente suspeitar. Assim, a teoria do Nível A é, pelo menos em um grau significativo, também uma teoria de níveis B e C. (WEAVER, 1949, p. 3).

O nível A envolve escolha, organização, codificação e transmissão da mensagem. Os problemas associados a essas etapas são, portanto, importantes também para os estudos do ciclo informacional, pois têm reflexos no armazenamento, na indexação e na recuperação da informação.

O ruído, por exemplo, é definido como uma perturbação que afeta o sinal transmitido, em qualquer das etapas do processo, de tal forma que a informação recebida não seja exatamente aquela que fora enviada. Um pouco de redundância permite que se corrija parte desse erro, de modo a aproximar a mensagem recebida da enviada:

Por exemplo, em um canal telegráfico sem ruído se poderia economizar cerca de 50% do tempo pela codificação adequada de mensagens. Isso não é feito e a maior parte da redundância de inglês permanece nos símbolos do canal. Isto tem a vantagem, no entanto, de permitir considerável ruído no canal. Uma fração importante de letras pode ser recebida incorretamente e ainda ser reconstruída pelo contexto. (SHANNON, 1948, p. 24).

Ruídos comuns são as distorções sonoras nas ligações telefônicas e as interferências nas transmissões de televisão por satélite. Mas em muitos casos as mensagens podem ser compreendidas, apesar dos ruídos, justamente porque há conteúdo redundante. Em outras palavras, os elementos do nível A da

comunicação (redundância e ruído) têm impacto direto nos elementos dos níveis B e C.

A capacidade de um canal de comunicação é uma medida da quantidade de informação que pode transitar por ele a cada unidade de tempo. A informação frequentemente não estará em seu estado original, mas terá tido sua forma alterada justamente para que possa ser transmitida. A essa alteração dá-se o nome de codificação.

Tome-se o exemplo de um pensamento. Ele não é transmitido diretamente, mas codificado em palavras (ou em um desenho, ou em sons), e essas palavras é que são comunicadas (de forma falada ou escrita). A codificação, portanto, é uma etapa crítica do processo de comunicação. Para cada canal de comunicação há uma codificação mais apropriada – transmitir um a ideia por texto requer elementos diferentes da transmissão da mesma ideia por imagens. A ideia é a mesma, o que difere é a codificação, para compatibilizar a informação a ser transmitida com o canal de comunicação a ser utilizado.

Weaver alerta para o fato de que o aperfeiçoamento da codificação normalmente implica aumentar o tempo necessário dessa etapa – codificações melhores são mais demoradas. Na prática, portanto, há um compromisso a se estabelecer entre o aproveitamento ideal do canal e a velocidade da comunicação.

Os conceitos discutidos quando da abordagem técnica – ruído, redundância, canal de comunicação, capacidade, codificação – também estão presentes na comunicação humana. Isso torna possível partir do trabalho de Shannon e Weaver e realizar estudos de sistemas de comunicação centrados em significado e usuários.

4 SIMILARIDADE COM O CICLO DOCUMENTÁRIO

Informação, mensagem, significado, semântica e receptor/usuário são termos usados na ciência da informação e que também estão presentes na teoria matemática da comunicação, mas não necessariamente representando os mesmos fenômenos. “[...] A ciência da informação há muito discute o conceito de informação [...]” (ZINS, 2008). Na teoria matemática da comunicação, informação é uma medida da liberdade de escolha que se tem ao se selecionar a mensagem que será transmitida. A informação

está relacionada nem tanto ao que diz, mas ao que se pode dizer WEAVER, e é explicitamente diferenciada de significado.

A palavra informação, nessa teoria, é usada em um sentido especial que não deve ser confundido com seu uso comum. Em particular, a informação não deve ser confundida com significado. Na verdade, duas mensagens, uma das quais fortemente carregada de significado e outra que seja puro disparate, podem ser exatamente equivalentes, do presente ponto de vista, em matéria de informação. (WEAVER, 1949, p. 4).

Esse é o modo técnico de abordar o problema. Quando se quer estudar os demais níveis de comunicação, o significado passa a ter importância. Convém lembrar a distinção entre definição e conceito proposta por Belkin (1978, p. 58) “[...] uma definição presumivelmente diz o que o fenômeno definido é, enquanto que um conceito é um modo de olhar para, ou de interpretar o fenômeno [...]”. Shannon (1948) e Weaver (1949) trabalharam com um conceito de informação que lhes foi útil, sem pretensão de criarem definições:

O conceito de informação não se aplica às mensagens individuais (como o conceito de significado), mas sim à situação como um todo, a unidade de informação indicando que nessa situação a pessoa tem uma quantidade de liberdade de escolha, na seleção de um mensagem, que é conveniente considerar como um padrão ou unidade de valor. (WEAVER, 1949, p. 4).

Como se vê, o conceito de informação usado na teoria matemática da comunicação é específico e estendê-lo a outras situações requer prudência – afinal, não é uma definição.

Brookes (1980) estabelece relação entre o que chama de informação objetiva e de conhecimento objetivo. O significado está no conhecimento, que requer interpretação da informação. Referindo-se à informação que comanda uma máquina, ele afirma que a informação que ela usa é simplesmente uma sequência programada de sinais; não foi estruturada em conhecimento, continua sendo informação objetiva. A informação objetiva, portanto, pode ou não ter significado, o que é compatível com o que foi definido na teoria matemática da comunicação.

No contexto da ciência da informação, o ciclo documentário congrega a produção, reunião, processamento técnico, disseminação e uso de documentos. De um modo mais completo e detalhado o ciclo documentário pode ser considerado um sistema, pois reúne em sua estrutura as três etapas centrais de qualquer sistema: entrada, processamento e saída. As etapas do ciclo documentário preparam as etapas subseqüentes em um movimento de retroalimentação constante, observando sempre o requisito do usuário, ou seja, o sistema deve estar calcado na perspectiva do atendimento das demandas informacionais dos usuários/clientes.

Segundo Robredo (2005), as doze operações ou etapas da cadeia ou ciclo documentário são:

- **Seleção:** escolha, dentre um conjunto de documentos, daqueles que podem responder às disciplinas ou tópicos cobertos por um dado serviço de documentação. Estes documentos têm vários suportes, dentre eles livros, revistas, documentos de arquivo, mapas, filmes discos, coleções de dados, entre outros;
 - **Aquisição:** atividade administrativa para adoção de providências de compra, intercâmbio, doação, pagamento, controle de chegada. No caso de documentos arquivísticos, a incorporação ao acervo de uma instituição se dará após as etapas de diagnóstico e estabelecimento das idades documentais. De modo amplo, esta etapa do ciclo documentário visa assegurar a incorporação dos documentos aos acervos;
 - **Registro ou tombamento:** atividade administrativa na qual são atribuídos números de registro para controle da incorporação dos documentos aos acervos das unidades informacionais. Esta atividade é de suma importância tanto para acervos resultados de compra ou doação, bem como para acervos arquivísticos. Nesta etapa fica consubstanciada a identificação da existência dos documentos;
 - **Descrição:** atividade de processamento técnico relacionada com o registro das características descritivas dos documentos.
- Esta providência é amplamente apoiada pelo diagnóstico situacional das condições físicas dos documentos. No caso de documentos arquivísticos, esta etapa propicia a compreensão dos princípios que regem os documentos, assim como as condições físicas das massas documentais acumuladas;
- **Análise ou condensação:** etapa de processamento técnico em que são desenvolvidas atividades de sumarização do conteúdo dos documentos, por meio de um número limitado de sentenças ou frases que devem expressar a essência temática do documento. Esta atividade, segundo Araújo Jr. (2007), é denominada de representação do conteúdo dos documentos e é realizada com o apoio de instrumentos de controle terminológico baseados em linguagem documentária;
 - **Indexação:** atividade de processamento técnico destinada à identificação dos conceitos de que trata o documento, expressando-os na terminologia utilizada pelo profissional da informação em linguagem natural ou com a ajuda de vocábulos ou termos derivados de linguagens documentárias;
 - **Armazenamento dos documentos:** etapa onde os documentos, a partir de dados que os descrevam por meio de sistemas de classificação, são armazenados de forma a serem encontrados quando solicitados pelos usuários nas unidades informacionais. Cada uma destas unidades possui especificidades de armazenamento em estantes, arquivos, microfichas, suportes magnéticos, entre outros;
 - **Armazenamento da representação condensada dos documentos:** etapa onde os elementos descritivos do conteúdo dos documentos, bem como os números que lhes foram atribuídos na etapa de registro ou tombamento, constituem um conjunto de dados que caracteriza, de forma condensada, cada documento. Estes elementos devem ser registrados ou inseridos em bases de dados para sua conservação, processamento e recuperação;
 - **Processamento da informação condensada:** atividade em que os registros

que contêm os elementos representativos dos documentos são submetidos a diversos processamentos manuais ou automatizados, que possibilitam a obtenção de vários produtos, nos quais a informação encontra-se reordenada e organizada de maneira a facilitar sua recuperação;

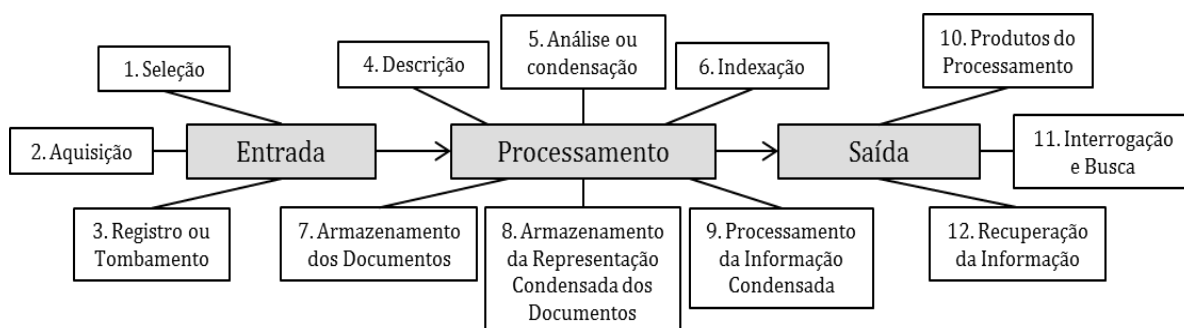
- **Produtos do processamento:** etapa onde são formados os produtos que vão compor o portfólio das unidades informacionais. Vários podem ser os serviços documentários a serem ofertados para os usuários. Podem-se destacar catálogos de diversos tipos, guias de fundos, cadastros, diretórios impressos ou eletrônicos, repositórios institucionais, portais de informação, bases de dados e de conhecimento entre outros;
- **Interrogação e busca:** etapa na qual a interação entre a gestão do ciclo documentário e a realização de pesquisa nos acervos se dá por uma ação de comunicação e mediação entre os profissionais da informação e os usuários da unidade informacional onde o ciclo documentário se desenvolve. No âmbito das unidades informacionais, esta atividade é conhecida como serviço de referência e se reveste de grande importância para a recuperação da informação; e

- **Recuperação da informação:** esta etapa do ciclo documentário constitui a finalidade do trabalho documentário. Ela é o resultado das operações realizadas no processo de busca. Segundo Araújo Jr. (2007), a recuperação da informação é reconhecida como a recuperação de referências de documentos em resposta às solicitações (demandas expressas por informação).

A figura 2 apresenta uma visão geral das operações/etapas do ciclo documentário associadas à entrada, processamento e saída elementos típicos da teoria geral de sistemas.

Considerando que uma das mais comuns formas de codificação de ideias é a palavra, então um documento pode ser visto como um canal de comunicação - e, de forma similar, uma pintura, um desenho, uma gravação de áudio, etc. Desse modo, o ciclo documentário tem similaridades com o um sistema de comunicação, sobretudo no que se refere às etapas finais do ciclo, ou seja, processamento da informação condensada, produtos do processamento, interrogação e busca e recuperação da informação. Além disso, é possível estudar esse sistema a partir do referencial da análise do problema técnico da comunicação, evidentemente com as adaptações conceituais pertinentes.

Figura 2 – Etapas do ciclo documentário



Fonte: Elaborado pelos autores, 2013.

Como em toda analogia, é preciso deixar claros os limites e as diferenças com o trabalho de referência. Quando a informação tem significado, trata-se dos níveis B e C definidos por Weaver, não do nível A.

A teoria matemática da comunicação trata da transmissão de bits. Já em um documento, o que se quer transmitir são ideias, significados. O conteúdo de um documento não é a materialização exata das ideias, mas o resultado da codificação delas em palavras, em vídeo, etc. Para cada meio de transmissão haverá uma codificação mais eficiente, mas as ideias essenciais devem estar lá, porque são elas que se quer fazer chegar ao destinatário.

Para que a comunicação seja bem sucedida, o destinatário e o emissor devem utilizar o mesmo esquema de codificação, de modo que um entenda o que o outro quis transmitir. Na elaboração de um texto, por exemplo, o adequado domínio do idioma e uma construção coerente e coesa facilitam sua compreensão. Nesse sentido, ambiguidades funcionam como um ruído na comunicação, pois permitem ao destinatário ter entendimento diferente da intenção do emissor da mensagem.

Na comunicação humana, o canal de comunicação é o conjunto de capacidades cognitivas dos agentes envolvidos. De um lado há o limite da capacidade de transmitir informações do emissor – sua habilidade para traduzir as ideias em um raciocínio articulado, sua fluência no idioma e no assunto; do outro lado há o limite do destinatário para absorver as informações, decodificar a mensagem e compreender o significado que ela contém. Ou seja, o limite de informação que pode transitar pelo canal depende da qualidade da codificação, que é afetada pelo meio de transmissão a ser usado e pelo domínio da linguagem específica apropriada ao assunto.

Não é demais frisar que a codificação deve estar otimizada para o meio de transmissão: uma cena ficará mais bem representada por uma fotografia ou por um desenho do que por um texto; alguns tipos de dados ficam mais inteligíveis se apresentados em uma tabela, ao invés de serem descritos em palavras. O papel é o suporte, o meio de transmissão das ideias.

A compreensão dos significados que o emissor pretendeu transmitir dependerá também da capacidade de decodificação do destinatário,

o usuário que receberá aquele documento e terá que interpretar o texto e extrair dele as informações. O sucesso dessa comunicação dependerá do conhecimento comum da linguagem, de forma que o que uma pessoa codificou (escreveu) a outra consiga decodificar (interpretar). O tempo disponível, o conhecimento do assunto, o domínio do idioma e a disposição física são fatores limitadores desse canal de comunicação.

Portanto, o ruído que pode afetar a mensagem, a redundância de seu conteúdo e as limitações do canal de comunicação devem ser avaliados também sob a perspectiva de suas consequências para a compreensão, pelo destinatário, dos significados que o emissor quis lhe transmitir.

Mas, qual o propósito de se estudar o ciclo documentário tendo por base uma teoria matemática? A ciência da informação tem característica multidisciplinar e oferece aos pesquisadores possibilidades de buscar em outras áreas de conhecimento elementos que possam contribuir para sua evolução.

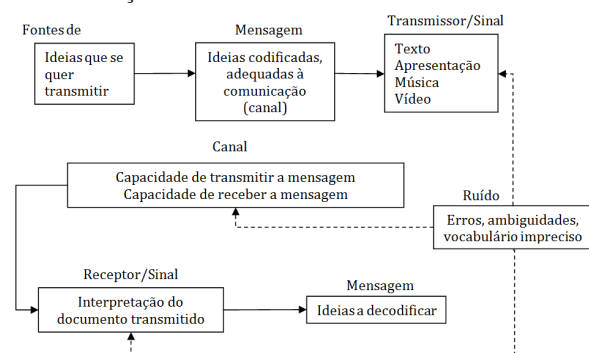
Assim, tanto a partir de um ponto de vista teórico quanto prático, cientistas da informação estão interessados na estrutura de seus objetos de estudo – informação. Mas, como os exemplos acima indicam, muitos cientistas sociais e comportamentais estão interessados em estruturas subjacentes também. Muitos engenheiros, baseados no trabalho de Shannon e Weaver, dentre outros, estão interessados em informações. O que, então, é distintivo sobre a teoria da ciência da informação? Estamos interessados em informação como um fenômeno social e psicológico. A informação que nós estudamos geralmente se origina de uma ação humana, de alguma forma, seja um feixe de dados emitido de um satélite ou o texto de um livro sobre a filosofia de Immanuel Kant. Nosso foco principal, mas não único, é em informação registrada e relacionamento das pessoas com ela. (BATES, 1999, p. 1047-1048).

Os principais conceitos tratados no modelo de Shannon e Weaver estão presentes na comunicação humana e é possível aproveitar parte do que foi desenvolvido na teoria matemática da comunicação para estudar o ciclo documental. As mensagens humanas, as interpretações e os significados

estão sujeitos aos mesmos fenômenos – ruído, redundância, restrição de canal. Há elementos essenciais a serem transmitidos, há codificação e decodificação. A explicitação de todos esses elementos amplia a compreensão de todo o processo de comunicação e pode contribuir para aprimorar tanto a produção de documentos de arquivo quanto as estratégias de armazenamento, busca e recuperação da informação.

Uma analogia entre o ciclo documentário e o esquema de comunicação descrito por Shannon (1948) está representada na Figura 3.

Figura 3 – Analogia entre um sistema de comunicação e o ciclo documental.



Fonte: Shannon (1948)

A mensagem é um produto intelectual, é a materialização das ideias do emissor – o resultado do processo de codificação –, e pode ser apresentada em diversas formas, cada qual adequada a um tipo de comunicação. Diferentes tipos de ruídos podem afetar a mensagem, desde os ruídos do tipo técnico – falhas ou perturbações em sua transmissão – até ruídos no próprio processo de codificação – distorções no registro das ideias, que afetarão a capacidade de compreensão do que se pretendeu comunicar.

Idealmente, o destinatário deve ser capaz de extrair do documento exatamente as ideias que o emissor quis transmitir. Por definição, toda diferença entre o que se tentou exprimir e o que o usuário compreendeu pode ser ruído ou perda. Será ruído se o usuário recebeu (entendeu, interpretou) algo que não foi enviado pelo emissor, e será perda se algo que foi transmitido não tiver sido recebido. Note-se que o ruído e a perda dependem, em parte, do destinatário da mensagem.

Se o destinatário tiver um entendimento incorreto sobre uma palavra em um texto, ele atribuirá a ela um significado diferente daquele usado pelo emissor, e sua interpretação do texto será outra. Terá, então, havido uma perda (o significado que o emissor codificou) em decorrência de um ruído (um sentido diferente para aquele texto, introduzido pela interpretação do destinatário).

Na analogia ilustrada na Figura 3, o canal de comunicação é formado pelas capacidades de transmitir e receber as informações. Ou seja, é o conjunto das capacidades técnicas de transmissão dos documentos, das características cognitivas do emissor e do receptor, de seus níveis de concentração e de interesse e ainda do tempo disponível para a transmissão.

5 CONCLUSÕES

Diversas etapas do ciclo documentário podem ser beneficiadas, em termos de eficiência e efetividade, se forem levadas em consideração características e restrições discutidas na teoria matemática da comunicação. A seleção, a aquisição e o tombamento de documentos devem guardar coerência com o serviço de informação que se pretende prestar, e isso implica considerar o canal de comunicação – como os usuários vão solicitar e receber as informações que buscam.

A descrição, a condensação e a indexação devem avaliar os efeitos da redundância presente nos documentos, e também o ruído da linguagem. Isso ajuda a dimensionar o próprio arquivo, facilita o armazenamento e torna a recuperação mais objetiva. É fundamental, porém, que se mantenha a neutralidade nessas etapas, cuidando-se para não inserir juízos de valor sobre o conteúdo que se quer preservar.

O processamento da informação condensada será mais eficiente se o armazenamento tiver sido de melhor qualidade, e os produtos deverão ser apresentados na linguagem adequada ao usuário (o que equivale à codificação apropriada para cada destinatário).

A atenção a esses elementos – ruído, redundância, canal de comunicação, codificação – tem reflexos positivos na etapa de interrogação e busca, porque o arquivo está em boa forma, organizado e livre de informações desnecessárias ou distorcidas, e acaba por facilitar a recuperação da informação, que é a própria finalidade do trabalho documentário.

STUDY OF THE SIMILARITIES BETWEEN THE MATHEMATICAL THEORY OF COMMUNICATION AND CYCLE DOCUMENTARY

ABSTRACT

This study, in the context of information science, addresses the mathematical theory of communication proposed by Claude Shannon in 1948 and its similarities with the cycle documentary regarding the stages ranging from selection, storage and processing technical documents, until the recovery of information retrieval by users. The theme is introduced by presenting four strands among the many communication systems approaches in information science, represented by: 1. scientific communication; 2. dichotomy between natural and documentary languages; 3. insertion of the mathematical theory of communication and into the information science 4. documentary cycle, which has in the information retrieval step the most significant one from the point of view of a communication system. Presents the definition, scope and characteristics of the mathematical theory of communication, strengthening its pioneering character to propose, under solid basis, the calculation of the amount of information that can be carried over a communication channel. Also addresses issues relating to neutrality of the communication channel that are present in discussions on information science, taking as an example the evaluation of documents in archival theory. Finally, are delimited the similarities of the mathematical theory of communication with the documentary cycle, where a document may be seen also as a communication channel, as well as any support that present information to be treated from the documentary point of view.

Key words:

Information Science. Mathematical theory of communication. Documentary chain. Communication systems.

Artigo recebido em 20/07/2013 e aceito para publicação em 19/07/2014

REFERÊNCIAS

ARAÚJO JUNIOR, R. H. de. **Precisão no processo de busca e recuperação da informação**. Brasília: Thesarus, 2007.

BATES, M. J. The invisible substrate of information science. **Journal of American Society of Information Science**, v. 50, n. 2, p. 1043-1050, 1999.

BELKIN, N. J. Information concepts for Information Science. **Journal of Documentation**, vol. 34, n. 1, p. 55-85, 1978.

BROOKES, B. C. The foundations of Information Science: Part I. Philosophical aspect. **Journal of Information Science**, vol. 2, n. 3-4, p. 125-133, 1980.

COLE, C. Calculating the information content of an information process for a domain expert using Shannon's Mathematical Theory of Communication: a preliminary analysis. **Information Processing & Management**, v. 33, n. 6, p. 715-726, 1997.

DURANTI, L. The concept of appraisal and archival theory. **The American Archivist**, vol. 57, n. 2, p. 328-344, Spring, 1994.

FIDALGO, A. Informação e Redundância. Os quadros da incerteza. In: SANTOS, J. M.; CORREIA, J. C. (Orgs). **Teorias da Comunicação**. Covilhã: Universidade da Beira Interior Editora, 2004. p. 15-28.

GUEDES, W. **Necessidades informacionais da auditoria pública: estudo qualitativo do fluxo de informação a partir da teoria matemática da comunicação**. 2013. Dissertação (Mestrado em Ciência da Informação) – Universidade de Brasília.

PERUCCHI, V. ARAÚJO JR. R. H de. Produção científica sobre inteligência competitiva da Faculdade de Ciência da Informação da Universidade de Brasília. **Perspectivas em Ciência da Informação**, v.17, n.2, p.37-56, abr./jun. 2012.

SHANNON, C. E. A Mathematical Theory of Communication. **The Bell System Technical**

Journal, v. 27, p. 379-423, 623-656, 1948.
Available in: <<http://cm.bell-labs.com/cm/ms/what/shannonday/shannon1948.pdf>>. Access:
03 jul. 2013.

WERSIG, G.; NEVELING, U. The phenomena of interest to information science. **The information Scientist**, v.9, n4, Dec. 1975.

WEAVER, W. **Recent contributions to The Mathematical Theory of Communication**, Sep 1949. Available in: <<http://ada.evergreen.edu/~arunc/texts/cybernetics/weaver.pdf>>. Access: 03 jul. 2013.