

UNIDADE 2

LÓGICA BOOLEANA APLICADA À RECUPERAÇÃO DA INFORMAÇÃO

2.1 OBJETIVO GERAL

Apresentar a álgebra booleana, seus conectivos e a maneira de utilizá-los como tradutores entre uma demanda por informação, realizada em linguagem natural, e um SRI, bem como seu uso nas diferentes interfaces de SRIs.

2.2 OBJETIVOS ESPECÍFICOS

Esperamos que, ao final desta Unidade, você seja capaz de:

- a) avaliar e identificar parâmetros de qualidade na recuperação da informação;
- b) diferenciar os conceitos de signo, termo e conceito, à luz da Semiótica;
- c) identificar termos e conceitos nas demandas por informação, de forma a obter resultados mais relevantes nas buscas em SRI;
- d) realizar análises conceituais de necessidades de informação e utilizar operadores booleanos para trabalhar com as inter-relações entre as características desejadas em um documento;
- e) converter/traduzir uma demanda por informação em linguagem natural para uma expressão booleana como estratégia de busca em um SRI.



2.3 NÃO FALAMOS A MESMA LÍNGUA

Mesmo quando nos comunicamos uns com os outros empregando a linguagem natural, frequentemente ocorrem mal-entendidos que nos obrigam a repetir, a usar sinônimos ou a explicar com mais detalhes exatamente o que queremos dizer. Nunca devemos consultar um SRI imaginando que ele é uma pessoa, ou que compreende nossa linguagem, embora muitos SRIs pareçam nos entender.

Infelizmente, em termos práticos, ainda não conseguimos nos comunicar com os SRIs usando linguagem natural. Então, a clareza das demandas a serem submetidas para que o SRI nos responda é uma questão-chave na recuperação de informações. Acontece, como já vimos na Unidade 1, que essas questões – ou demandas por informação – são pensadas ou formuladas, em um primeiro momento, em linguagem natural, mas precisamos traduzi-las para uma linguagem que o SRI compreenda.

Figura 12 – Para entender o que queremos, os SRIs, muitas vezes, precisam de informações mais específicas



Fonte: produção do próprio autor a partir de imagens da internet.⁴

Nesta Unidade, vamos trabalhar essa questão apresentando-lhe a *álgebra booleana*, uma linguagem “compreensível” pela maioria dos SRIs. Mas, antes que possamos traduzir uma demanda por informações da linguagem natural para a álgebra booleana, veremos algumas noções de análise conceitual. Elas nos permitirão analisar e superar os problemas da própria linguagem natural, sua ambiguidade e imprecisão com relação ao que exatamente queremos dizer em uma demanda por informações.

⁴ Primeira imagem: pessoa usando computador. Autor: *Joonspoon*. Disponível em: <https://en.wikipedia.org/wiki/Programmer#/media/File:Programmer_writing_code_with_Unit_Tests.jpg>; segunda imagem: *Batman*. Autor: *spadge6868*. Disponível em: <https://en.wikipedia.org/wiki/Programmer#/media/File:Programmer_writing_code_with_Unit_Tests.jpg>. Acesso em: 30 jul. 2021.

2.4 CONCEITOS QUE NOS AJUDAM A “CONVERSAR” COM UM SRI

Pense nas seguintes demandas por informação (Figura 13):

Figura 13 – Demandas por informação

Demandas de um usuário	“Desejo manuais sobre técnicas de redação de documentos administrativos.”
	“Preciso recuperar artigos científicos sobre prevenção da dengue.”
	“Quero relações de questões para concursos públicos para a área de assistente social.”

Fonte: produção do próprio autor.

Como já foi visto na Unidade 1, um SRI é um sistema que funciona como intermediário entre as demandas de informação formuladas por um usuário e uma coleção de registros de informação. Antes, as necessidades do usuário poderiam ser buscadas na seção de referência de uma biblioteca, cujo bibliotecário indicaria matérias e fontes de informação ali disponíveis sobre essas questões.

No entanto, hoje, é cada vez mais comum que catálogos de bibliotecas, bases de dados e outras fontes de informação estejam disponíveis diretamente na *web*. Nesses casos, o usuário não terá a assistência de um bibliotecário de referência com conhecimentos do acervo da biblioteca e capaz de orientar com precisão se determinado material está disponível ou não.

Repare também que as demandas por informação, no caso de uma pesquisa feita com o bibliotecário, na biblioteca, foram formuladas usando a linguagem natural, tendo sido, portanto, compreendidas pelo profissional de referência. Caso ele não compreenda exatamente uma questão demandada por um usuário, pode solicitar-lhe que esclareça melhor sua questão, e que diga quais os aspectos específicos de sua demanda.



Multimídia

A comunicação humana (Figura 14) realizada através de um diálogo em linguagem natural é um fenômeno cognitivo muitíssimo complexo, específico da espécie humana, sobre cuja origem existem várias teorias.

Para saber um pouco mais sobre a linguagem humana, esse tema fascinante, leia o artigo intitulado “Da linguagem humana ao processamento humano”, de Isabel Hub Faria, publicado na *Revista Veredas*, uma revista de Estudos Linguísticos da *Universidade Federal de Juiz de Fora*. Ele pode ser encontrado em: <<http://www.ufjf.br/revistaveredas/files/2009/12/artigo015.pdf>>. Acesso em: 30 jul. 2021.

Figura 14 – Comunicação



Fonte: Flickr.⁵

No contexto de recuperação de informações, os usuários devem se comunicar com o SRI especificando suas necessidades. A especificação das demandas por informação de um usuário é expressa, num primeiro momento, através da *linguagem natural*, que é a que utilizamos no nosso dia a dia.

A linguagem natural é estudada pela disciplina científica chamada Linguística, segundo a qual essa linguagem compreende as seguintes dimensões: sintaxe, semântica e pragmática.

A sintaxe trata das regras para formar sentenças corretas numa linguagem. “O menino chutou a bola.” é uma sentença sintaticamente correta na língua portuguesa; já “Chutou o menino bola a.” é uma sentença incorreta.

Por sua vez, a semântica trata do significado correto das sentenças de uma linguagem. “O menino chutou a bola.” é uma sentença semanticamente correta na língua portuguesa; já “A bola chutou o menino.”, não, uma vez que o que a sentença afirma não corresponde à realidade: uma bola é incapaz de chutar um menino (a não ser que estejamos falando de sonhos, ou de uma realidade absurda).

Por fim, a pragmática trata do uso real da linguagem no contexto da comunicação direta entre seres humanos. Nesses contextos práticos (ou pragmáticos), muitas vezes, ao falarmos uma palavra, queremos dizer exatamente o contrário. Por exemplo, quando uma mãe quer negar algo ao filho, ela fala “Pois sim!”. Veja que o que ela está querendo dizer, na verdade, é “não”. Outro exemplo seria a frase “Vi as montanhas voando

⁵ Autor: *Oficial GDC*. Disponível em: <<https://www.flickr.com/photos/officialgdc/5058378176>>. Acesso em: 30 jul. 2021.

sobre a França.”, uma frase ambígua em termos semânticos, mas que pode ser resolvida pragmaticamente, pois parece óbvio que é o falante, e não as montanhas, quem está voando.

A semântica é o ramo da linguística que mais nos interessa aqui, uma vez que queremos comunicar nossas demandas por informação aos SRIs com um significado preciso.



Atenção

Semântica é o ramo da Linguística que trata do significado. As expressões da nossa linguagem *significam* alguma coisa, e isso é o que nos permite a comunicação. (CHIERCHIA, 2003).

“Significar”, nesse contexto, é *referir-se* a algum fenômeno existente. A linguagem é o instrumento para a comunicação mútua de descrições ou de significados comuns, das coisas do mundo que nos cerca, para podermos agir cooperativamente a partir de uma visão do universo que compartilhamos. Mas nem tudo é simples e direto na comunicação humana através da linguagem. Palavras de uma linguagem podem ter vários significados, por exemplo: tênis, que pode designar tanto um tipo de calçado quanto um tipo de esporte; porca, que pode designar uma peça mecânica, que funciona juntamente com um parafuso, para afixar outras peças, ou um animal, e manga, que pode designar uma fruta ou uma parte das camisas. Palavras diferentes também podem ter o mesmo significado, como carro e automóvel, ou ainda aipim, macaxeira e mandioca.

Nos casos das fontes de informação na *web*, onde não existe um bibliotecário de referência, é preciso que se tenha a maior precisão possível na especificação das demandas por informação de um usuário. Isso porque ainda estamos longe de termos SRIs que compreendam a linguagem humana. Se um usuário não pode falar com um SRI, ele, no mínimo, pensou nessas questões, formulou mentalmente suas demandas por informação. Precisamos, então, de conhecimento e técnicas sobre como tornar compreensíveis a um SRI nossas demandas por informação, que foram formuladas ou pensadas em linguagem natural. Como dito anteriormente, esse é o objetivo desta Unidade.



2.4.1 Atividade

Esta é uma atividade prática e, para realizá-la, você precisará de um dispositivo (computador, *tablet*, celular etc.) com acesso à internet.

Imagine que um gestor da área de Saúde, o Secretário de Saúde de um município, queira prevenir e diminuir os acidentes de trânsito em sua cidade. Ele está interessado em saber as estatísticas

das sequelas e incapacidades causadas por acidentes de trânsito. Vamos simular a sua busca por informações?

Então, vá ao *Google* (<<https://www.google.com.br/>>) e formule duas consultas:

- 1ª consulta: estatísticas de sequelas e incapacidades devidas a acidentes de trânsito;
- 2ª consulta: “estatísticas de sequelas e incapacidades” e “acidentes de trânsito”.

É melhor que você abra o *Google* em duas janelas (abas) diferentes e formule cada consulta em uma delas.

Avalie e compare os resultados, depois responda às questões propostas.

- a) O número de resultados encontrados pelo *Google* foi o mesmo nas duas questões? Explique os resultados encontrados.

- b) Qual das duas consultas trouxe resultados mais relevantes? Por quê?

Resposta comentada

- a) A consulta sem aspas (1ª consulta) recupera um número maior de resultados que a consulta com aspas (2ª consulta). Nas consultas que eu fiz, encontrei 183.000 resultados na primeira, contra 340 na consulta com aspas. Percebi também que, conforme examinamos os resultados das últimas páginas apresentadas pelo *Google*, eles se tornam cada vez menos relevantes para as necessidades de informação de nosso Secretário municipal de Saúde;
- b) os resultados da consulta utilizando aspas, por outro lado, são bastante relevantes e trazem grande potencial de atendimento às demandas do nosso Secretário. Isso acontece porque, com as aspas, o *Google* traz resultados em que as palavras estão juntas e na mesma sequência que a da consulta submetida. Por exemplo, em “acidentes de trânsito” – com as palavras nessa ordem exata –, temos um significado distinto de “acidentes de trabalho” ou “acidentes vasculares cerebrais”. Nesses casos, temos, entre aspas, três palavras, mas cada um desses agrupamentos designa uma coisa diferente.

Aproveitando o exemplo da busca por “acidentes de trânsito” (dado na atividade anterior), agora veremos uma questão que será interessante sempre que você estiver utilizando um SRI: a avaliação dele como usuário.

Existem métricas de avaliação consagradas, que são chamadas de *revocação* e *precisão*. Elas serão conceitualizadas a seguir, juntamente com



relevância, um conceito que, por entrar em ambas as definições, será abordado primeiramente.

A *relevância* é definida por Saracevic (2007) como uma relação entre a informação, ou os objetos informacionais recuperados de um SRI (por exemplo, documentos), e as necessidades de informação dos usuários. A relevância expressa o quanto os usuários de um SRI avaliam que os resultados recuperados serão úteis para suas necessidades, sendo, portanto, uma medida dessa relação.

A partir do conceito de relevância, podemos definir, então, precisão e revocação.

Precisão é a razão entre o número de registros relevantes recuperados por dada consulta e o número de registros recuperados nessa mesma consulta (Figura 15).

Figura 15 – Precisão

Na forma de uma equação:

$$P = \frac{\text{n.º de registros relevantes recuperados}}{\text{n.º de registros recuperados}}$$

Fonte: produção do próprio autor.

Já *revocação* é a razão entre o número de registros relevantes recuperados por uma consulta e o número total de registros relevantes existentes na base de dados (Figura 16).

Figura 16 – Revocação

Na forma de uma equação:

$$R = \frac{\text{n.º de registros relevantes recuperados}}{\text{n.º total de registros relevantes}}$$

Fonte: produção do próprio autor.

Embora possamos calcular a precisão de uma consulta, não o podemos fazer com a revocação, a menos que saibamos, de antemão, quantos documentos relevantes para uma consulta existem em determinada base de dados.

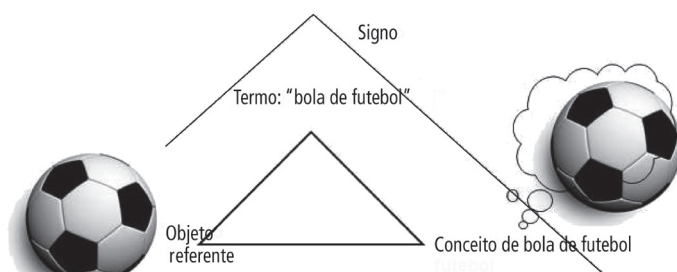
A revocação e a precisão foram estabelecidas como métricas nas primeiras experiências com um SRI, na década de 1960, em um ambiente de laboratório totalmente controlado e que apresentava as seguintes características:

- a) os experimentos foram realizados utilizando-se uma coleção de documentos conhecida;
- b) foi realizado um conjunto determinado de consultas;
- c) as avaliações foram feitas por especialistas, que determinaram quais documentos seriam relevantes para uma consulta. Isso significa que a relevância foi utilizada como atributo de documentos em relação a consultas específicas.

As expressões *acidente*, *acidente de trânsito* e *acidentes vasculares cerebrais* são formadas, respectivamente, por uma, duas e três palavras que *designam uma única coisa ou objeto no mundo*. Chamamos de *conceito*, que é uma unidade de pensamento, aquilo que designa cada uma das expressões. Como, na comunicação através da linguagem, não conseguimos ler os pensamentos uns dos outros, os conceitos precisam ter uma manifestação linguística, a que chamamos de *termo*. O termo, por sua vez, é uma unidade linguística de significado; *um termo significa uma (única) coisa*. Uma característica do conceito/termo é a *monorreferencialidade*, ou seja, o fato de referir-se sempre a *um único objeto no mundo*.

Um conceito é, portanto, uma relação entre três elementos: o pensamento que ele provoca ou desencadeia, um objeto do mundo, ao qual esse pensamento se refere, e o termo – uma entidade linguística –, que faz a mediação entre o pensamento e o objeto. Essa relação entre os três elementos foi proposta pelo filósofo americano *Charles Sanders Peirce* e é a base da Semiótica, a disciplina que estuda todos os sistemas de signos, ou seja, as coisas que significam algo. Essa relação pode ser ilustrada pelo triângulo semiótico, ilustrado na Figura 17:

Figura 17 – Triângulo semiótico usando como exemplo o objeto *bola* e o termo/conceito *bola de futebol*



Fonte: produção do próprio autor.

Para ilustrar essas ideias, seguem-se vários exemplos de conceitos/termos:

- a) bola;
- b) bola de futebol;
- c) atestado de bons antecedentes;
- d) atestado médico;
- e) atestado de óbito;
- f) homem;
- g) homem de negócios;
- h) engenharia;
- i) engenharia de produção;
- j) engenharia de produção de alimentos.

Todos os conceitos/termos listados anteriormente designam ou significam coisas diferentes, embora possam ter palavras em comum.

Agora que possuímos as noções de conceito e termo, vejamos o que podemos concluir em relação a uma tentativa de recuperar informações no *Google*. Imagine que fizemos uma busca sobre "estatísticas de seque-

las e incapacidades causadas por acidentes de trânsito”. Se você fez a Atividade 1, vai ficar mais fácil acompanhar esse raciocínio.

Para ajudar, pode ser útil consultar o próprio *Google*, para que ele indique explicações sobre como indexa as milhões de páginas da *web* (o *Google* cria índices de palavras encontradas nas diferentes páginas disponíveis na *web*).



Multimídia

Figura 18 – Discovery



Discovery SEO: Como o Google indexa a Internet?

Fonte: indisponível.

Se você quer saber mais sobre como o processo de indexação automática das páginas *web* é feito pelo *Google*, pode encontrar mais informações em: <<http://googlediscovery.com/2010/06/30/discovery-seo-como-o-google-indexa-a-internet/>>. Acesso em: 30 jul. 2021.

Podemos concluir que, se quisermos obter resultados mais relevantes e pertinentes de um SRI, devemos ter um meio de passar para ele termos (e não simplesmente palavras isoladas) que expressem nossas necessidades de informação. Assim, devemos ser capazes de identificar, na formulação de nossas demandas por informação em linguagem natural (ou nas de um usuário), *quais conceitos/termos existem* e *quais unidades de significado* ou *referência* a objetos únicos do mundo existem, numa expressão linguística em linguagem natural. Essa atividade é chamada de *análise conceitual*.



2.4.2 Atividade

Analise as afirmativas a seguir e identifique quais dos conceitos envolvidos na qualidade da recuperação da informação (revocação, precisão e relevância) cada uma representa.

- a) Foram recuperados 40 documentos relevantes sobre o tema *nutrição esportiva*, de um total de 150 documentos que são relevantes para a área em questão.

- b) De um total de 1.000 documentos recuperados sobre o tema *redação do Exame Nacional do Ensino Médio (ENEM)*, foram encontrados 50 documentos relevantes.

Resposta comentada

Respostas:

- a) revocação;
b) precisão.



2.4.3 Atividade

Observe os itens a seguir e identifique o que pode ser considerado um signo, um termo e um conceito. Explique a diferença entre esses elementos à luz dos estudos da Semiótica.

a)

Figura 19 – Aparelho telefônico



Fonte: Wikipédia.⁶

b) TELEFONE

c) Aparelho destinado a transmitir, a distância, a voz ou qualquer outro som, por meio do eletromagnetismo (com base na definição do dicionário *Michaelis*).

⁶ Autor: *Bran*. Disponível em: <https://pt.wikipedia.org/wiki/Telefone_de_disco#/media/File:Telefon04_2.jpg>. Acesso em: 30 jul. 2021.

Resposta comentada

- a) signo;
- b) termo;
- c) conceito.

No exemplo, vemos que a imagem do telefone é um sinal e representa um objetivo no mundo, uma única coisa. Esse objeto/coisa é definido pelo termo TELEFONE, que, por sua vez, pode ser definido pelo conceito expresso no item c).

2.5 ANÁLISE CONCEITUAL DE DEMANDAS POR INFORMAÇÃO EXPRESSAS EM LINGUAGEM NATURAL

Podemos resumir as questões envolvidas na análise conceitual (Figura 20) de uma demanda por informação feita em linguagem natural e a ser pesquisada num SRI da seguinte maneira:

Figura 20 – Análise conceitual

CONCEITO
<ul style="list-style-type: none">• Unidade mental (pensamento) de conhecimento, compartilhada por uma comunidade através de um termo linguístico. Se refere a um objeto único (monorreferencialidade). Exemplos: chá, café com leite etc.
TERMO (Expressão linguística do conceito)
<ul style="list-style-type: none">• Manifestação linguística de um conceito. Pode ser formado por uma ou mais palavras. Exemplos: chá, café com leite etc.
DEMANDA POR INFORMAÇÃO
<ul style="list-style-type: none">• Conjunto de termos expressando um assunto, uma necessidade de informação, uma questão de pesquisa. Exemplo: “valor nutricional do café com leite no desjejum”.

Fonte: produção do próprio autor.

Ao longo da história do pensamento humano, os problemas decorrentes da relação entre a linguagem natural e as coisas no mundo que ela designa foram objeto de reflexão por parte de diferentes pensadores. Como já vimos, esse estudo é chamado de Semântica.

A seguir, analisaremos contribuições de pensadores que se debruçaram sobre a questão que a Figura 21 anuncia:

Figura 21 – Problema



Fonte: produção do próprio autor a partir de imagem do Pixabay.⁷

O linguista americano *Charles Fillmore* afirmava que, para toda expressão em linguagem natural, haveria sempre alguns ou todos daquilo que ele chamava de *casos conceituais*: agente, objeto, instrumento, finalidade, produto, modo e lugar. Assim, por exemplo, na sentença: “João cortou a maçã com a faca, com presteza, para fazer a torta, ontem, na cozinha.”,

- a) João é o agente;
- b) maçã é o objeto;
- c) a faca é o instrumento;
- d) com presteza é o modo;
- e) na cozinha é o lugar.

O bibliotecário indiano *Shiyali Ramamrita Ranganathan* também aborda o problema propondo que, em um texto em linguagem natural – o título de um livro, por exemplo –, haveria sempre as seguintes categorias fundamentais: personalidade, matéria, energia, espaço e tempo, ou seja, coisas, atributos (das coisas), processos, espaço e tempo (RANGANATHAN, 1967). Por exemplo, no título: “Controle de doenças viróticas do caule do arroz no inverno de 1967 em Madras”,

- a) agricultura seria o assunto ou faceta básica;
- b) controle seria a manifestação da categoria *energia*;
- c) doenças viróticas seria a manifestação da categoria *matéria*;
- d) caule, bem como arroz, seriam a manifestação da categoria *personalidade*;

⁷ Autor: *McLac2000*. Disponível em: <<https://pixabay.com/pt/monumento-pensador-acho-que-vigor-1027558/>>. Acesso em: 30 jul. 2021.

- e) inverno, bem como 1967, seriam a manifestação da categoria *tempo*;
- f) Madras seria a manifestação da categoria *espaço*.

Ranganathan (1967) propunha que essa análise de categorias ou facetas servisse para criar uma notação que permitiria, então, armazenar e encontrar um livro nas estantes de uma biblioteca.

A partir das ideias de *Ranganathan*, o *Classification Research Group* (CRG) da Inglaterra, na década de 1950, desenvolveu outro conjunto de categorias fundamentais: coisas, tipos, partes, materiais, processos, operações, pacientes, produtos, subprodutos, agentes, espaço e tempo. O mesmo título do livro analisado segundo as categorias básicas de *Ranganathan* ficaria assim, de acordo com as categorias propostas pelo CRG:

- a) arroz seria a categoria *coisa*;
- b) caule seria a categoria *parte*;
- c) doenças viróticas seriam a categoria *paciente*;
- d) controle seria a categoria *operação*;
- e) Madras seria a categoria *espaço*;
- f) inverno de 1967 seria a categoria *tempo*.

A seguir, procurando responder à questão da análise conceitual, feita na figura anterior, daremos um conjunto de indicações (Figura 22):

Figura 22 – Que conceitos/termos existem em uma demanda por informação expressa em linguagem natural?

Em primeiro lugar, deve-se procurar o significado do conceito (termo) dentro do contexto da demanda por informação.

Um indicador de que uma palavra ou conjunto de palavras (termo) é um conceito é que este tem sempre um **único referente**, ou seja, indica sempre um **único objeto** no mundo, mesmo sendo composto por várias palavras; lembre-se de exemplos de conceitos/termos como: atestado de bons antecedentes, ataque cardíaco, homem de negócios etc.

Quando se desconfia de que mais de uma palavra do tema de pesquisa formam um único conceito, deve-se verificar se cada uma delas, separadamente, tem um significado **diferente** daquele de quando elas são coordenadas formando um único termo.

Em conceitos expressos por uma única palavra (termo, conceito), deve-se verificar o seu significado no contexto do tema de pesquisa.

Fonte: produção do próprio autor a partir de imagens da internet.⁸

Para que as indicações listadas anteriormente fiquem claras, vejamos o seguinte exemplo de demanda por informação: “cartórios que emitem

⁸ Primeira imagem: lupa. Autor: *ClkerFreeVectorImages*. Disponível em: <<https://pixabay.com/pt/ver-pesquisa-assuntos-s%C3%ADtio-web-42931/>>; segunda imagem: digital. Autor: *OpenClipartVectors*. Disponível em: <<https://pixabay.com/pt/impress%C3%A3o-digital-contato-crime-150159/>>; terceira imagem: diferente. Autor: *F. Delventhal*. Disponível em: <https://www.flickr.com/photos/krossbow/9984030564/>; quarta imagem: quebra-cabeças. Autor: *Willi Heidelberg*. Disponível em: <https://commons.wikimedia.org/wiki/User:Atlasowa#/media/File:Puzzle2_found_bw.jpg>. Acesso em: 30 jul. 2021.

atestado de óbito no município de Niterói” (Niterói é um município do estado do Rio de Janeiro):

- a) se pesquisarmos por “atestados”, isoladamente, poderemos obter como resultado: atestados de bons antecedentes, atestados médicos, atestados de residência etc.;
- b) se pesquisarmos por “óbito”, isoladamente, poderemos obter: número de óbitos, causas de óbitos etc.

Essa diversidade de resultados para as palavras isoladas indicam que *atestado de óbito* é um único conceito. Então, como fazer para que o SRI entenda que queremos as palavras juntas, formando um único conceito? É o que veremos a seguir.

2.6 CONECTIVOS BOOLEANOS

Seja a seguinte consulta (Figura 23) submetida a um SRI hipotético:

Figura 23 – Consulta

“Quero recuperar, no banco de dados de produtos disponíveis nesta rede de supermercados, chá verde das marcas ‘Leão’ ou ‘Chinezinho’ e que não sejam em sachê.”

Fonte: produção do próprio autor.

Se analisarmos essa demanda por informação em linguagem natural, observaremos que especificamos nossa necessidade listando um conjunto de *características obrigatórias* que a informação desejada deve ter, *características alternativas* (marcas “Leão” ou “Chinezinho”) e *características não desejáveis* (“que não sejam em sachê”).

Ao subtermos uma demanda por informações a um SRI, a ideia é que, quanto mais conseguirmos precisar o que desejamos, mais chance teremos de que o SRI recupere informações relevantes para nós. A maneira que temos de especificar a informação desejada é enumerarmos e detalharmos suas características. Assim, se nos dirigimos a uma biblioteca porque estamos precisando de um manual de redação de documentos administrativos para a administração federal, mais chance teremos de encontrar o manual desejado em uma coleção que poderia incluir manuais de redação jornalística, de redação de trabalhos acadêmicos, de redação de correspondência, de mecânica de automóveis etc. Veja a diferença entre as demandas:

- a) preciso de um manual;
- b) preciso de um manual de redação de documentos administrativos para a administração federal.



As características do documento que queremos obter na biblioteca como “manual”, “redação”, “documentos administrativos” e “administração federal” são expressas por meio de conceitos/termos. Quando se trata de documentos em uma biblioteca, a característica mais utilizada para descrevê-los é o seu assunto, mas outras, como forma (livros, revistas, dicionários) ou gênero (prosa, verso, literatura infantil, documentação técnica etc.) também são muito utilizadas.

Uma demanda por informação feita em linguagem natural é, então, um conjunto de conceitos/termos inter-relacionados, que expressam quais características da informação desejada são obrigatórias, quais são alternativas e quais não são desejadas. As relações entre as características de um documento podem ser dos seguintes tipos:

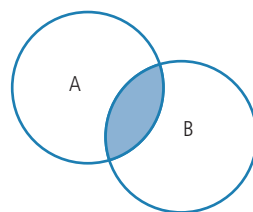
- a) simultaneidade de duas características: as duas características devem existir nos documentos, obrigatória e simultaneamente;
- b) *alternância de duas características*: são desejados documentos que tenham *uma ou outra* característica, *ou ambas*;
- c) *exclusão de características*: os documentos *não devem* possuir determinada característica.

Esses tipos de inter-relação entre as características desejadas nos documentos são feitos através do uso dos chamados operadores booleanos, que expressam com precisão essas relações.

A álgebra booleana é a parte da Matemática que trabalha com relações entre conjuntos. As operações booleanas básicas são a *intersecção* entre os elementos de dois conjuntos, cuja notação é \cap , a *união* entre os elementos de dois conjuntos, cuja notação é \cup , e a *negação* dos elementos de um conjunto, cuja notação é \neg . Os operadores booleanos são três:

- a) *E*: quando dois conceitos/termos são unidos pelo operador booleano *E*, isso significa que os documentos desejados devem possuir as características expressas pelos dois conceitos/termos, obrigatória e simultaneamente. O operador booleano *E* realiza a intersecção entre dois conjuntos *A* e *B* – $A \cap B$ –, sendo representado como na Figura 24:

Figura 24 – O conectivo booleano E, que representa a intersecção entre os conjuntos A e B, está representado graficamente pela região azul

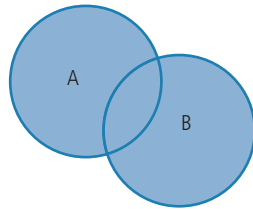


Fonte: produção do próprio autor.

- b) *OU*: quando dois conceitos/termos são unidos pelo operador booleano *OU*, isso significa que basta que os documentos possuam ou uma ou outra das características expressas pelos dois conceitos/termos; também são aceitos documentos que possuam ambas as características, já que estes atendem ao critério de possuírem pelo menos uma das duas características (possuem ambas). O operador

booleano OU realiza a união entre dois conjuntos A e B – $A \text{ OU } B$ –, sendo representado como na Figura 25:

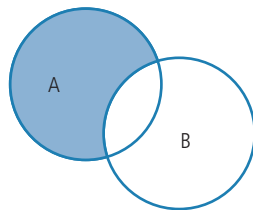
Figura 25 – O conectivo booleano OU, que representa a união dos conjuntos A e B, está representado graficamente por toda a região em azul



Fonte: produção do próprio autor.

- c) **NÃO**: quando um conceito/termo é precedido pelo operador booleano NÃO, isso significa que os documentos desejados não devem possuir a característica expressa pelo conceito/termo. O operador booleano NÃO realiza a diferença entre dois conjuntos A e B – $A \text{ NÃO } B$ –, ou seja, o conjunto de documentos que possui a característica A, mas não a característica B, sendo representado como na Figura 26:

Figura 26 – O conectivo booleano NÃO, que representa a diferença entre o conjunto A menos o B, está representado graficamente pela região que está em azul



Fonte: produção do próprio autor.

Mas como os conectivos funcionam em uma pesquisa? Vamos a um exemplo. Suponha os dois documentos listados a seguir (doc 1 e doc 2) e as palavras-chave que os descrevem:

- doc 1 – palavras-chave: AIDS, tratamento, criança;
- doc 2 – palavras-chave: AIDS, imunização, criança, índice de mortalidade.

Ao fazermos consultas utilizando os conectivos booleanos, recuperaríamos os seguintes documentos:

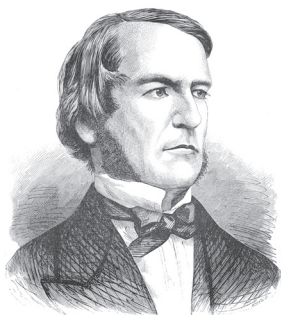
- consulta: **AIDS .E. criança** – doc 1, doc 2;
- consulta: **AIDS .E. tratamento** – doc 1;
- consulta: **AIDS .E. imunização** – doc 2;
- consulta: **AIDS .OU. imunização** – doc 1, doc 2;
- consulta: **AIDS .NÃO. índice de mortalidade** – doc 1.



Curiosidade

De onde vem a álgebra booleana?

Figura 27 – George Boole



Fonte: *Wikimedia Commons*.¹¹

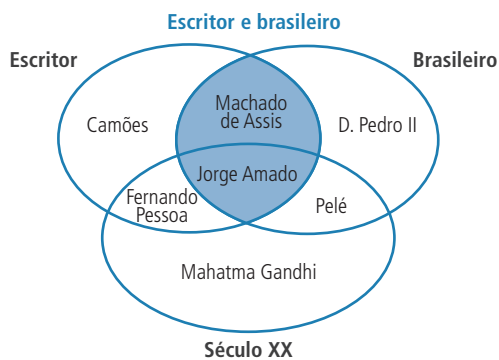
George Boole (1815-1864) (Figura 27) nasceu em Lincoln, Inglaterra. Era um gênio matemático que foi, em grande parte, autodidata. Sua nomeação como o primeiro professor de Matemática na *University College Cork*, na Irlanda, em 1849, proporcionou a ele a oportunidade de desenvolver seu trabalho mais importante, chamado *Investigation of the Laws of Thought* (em tradução livre: *Investigação sobre as leis do pensamento*).

Boole pode ser descrito como o pai da era da informação. A criação da álgebra booleana e da lógica simbólica fizeram dele um pioneiro da nova Matemática. Seu legado nos rodeia em todos os lugares no século XXI: nos computadores, no armazenamento e recuperação de informações, nos circuitos eletrônicos e controles que suportam a vida, na aprendizagem e nas comunicações.

Fonte: GEORGE, c2017.

Seguem-se, nas figuras de 28 a 33, outros exemplos de como são processados os três operadores booleanos:

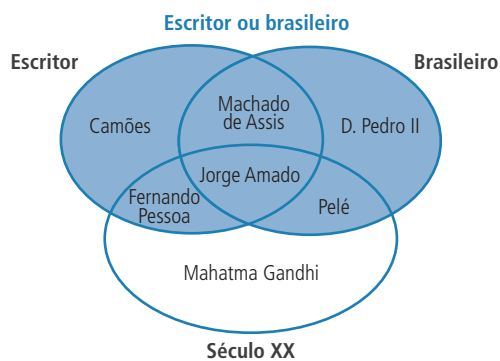
Figura 28 – Exemplo de conectivo booleano E



Fonte: produção do próprio autor.

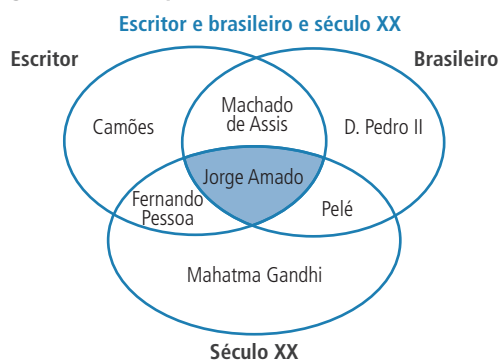
⁹ Em domínio público. Disponível em: <https://commons.wikimedia.org/wiki/File:PSM_V17_D740_George_Boole.jpg>. Acesso em: 30 jul. 2021.

Figura 29 – Exemplo de conectivo booleano OU



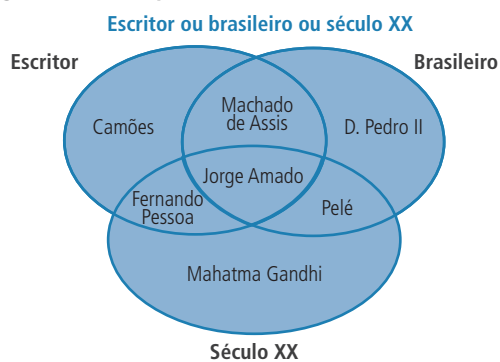
Fonte: produção do próprio autor.

Figura 30 – Exemplo de dois conectivos booleanos E



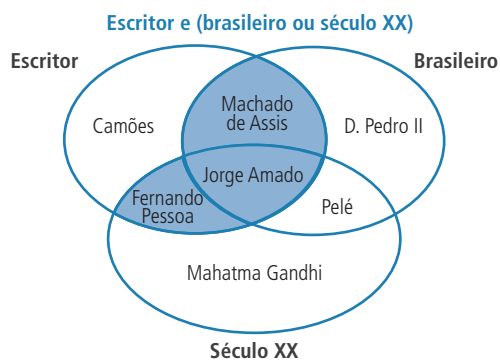
Fonte: produção do próprio autor.

Figura 31 – Exemplo de dois conectivos booleanos OU



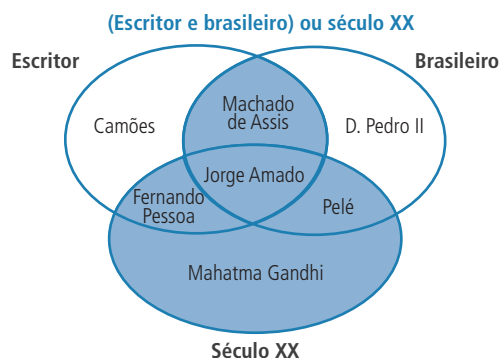
Fonte: produção do próprio autor.

Figura 32 – Exemplo de uso de parênteses com conectivos booleanos



Fonte: produção do próprio autor.

Figura 33 – Exemplo de uso de parênteses com conectivos booleanos



Fonte: produção do próprio autor.

Você percebeu que, nos exemplos das Figuras 32 e 33, algumas expressões booleanas aparecem entre parênteses? Os parênteses servem para priorizar os operadores booleanos que serão avaliados em primeiro lugar, de forma semelhante ao que fazemos quando executamos operações matemáticas (Figura 34).

Figura 34 – Exemplo

Por exemplo: $5 \times 3 + 4 = 19$, mas $5 \times (3 + 4) = 35$.

Fonte: produção do próprio autor.

Vamos a outro exemplo. Imagine que desejamos recuperar artigos sobre a economia nas regiões Norte e Nordeste. A expressão booleana *Economia E Norte OU Nordeste* trará um resultado muito distinto do obtido com a expressão booleana *Economia E (Norte OU Nordeste)*. Na primeira expressão, sem parênteses, o SRI recuperará documentos sobre economia e região Norte, ou seja, sobre a economia da região Norte, o que corresponde parcialmente ao que queremos. No entanto, ao prosseguir executando a expressão booleana, o SRI processará os resultados parciais já obtidos (documentos sobre economia da região Norte), fazendo um OU com *qualquer coisa da região Nordeste* (inclusive coisas relacionadas à economia, mas não somente, podendo também abranger turismo, cultura, demografia etc.).

Constatamos, assim, que a expressão entre parênteses – *Economia E (Norte OU Nordeste)* – é a que realmente manifesta a nossa demanda por informação.

Então, devemos e podemos, sempre que for necessário, empregar os parênteses nas expressões booleanas para traduzir exatamente o que pretendemos obter do SRI consultado.



2.6.1 Atividade

Imagine que você deseja fazer pesquisas utilizando um SRI similar ao *Google*. Analise as demandas a seguir e desenvolva, para cada uma delas, as expressões booleanas que utilizaria para encontrar informações relevantes para estas necessidades de informação:

a) artigos de revisão sobre política industrial dos estados da região Sudeste e Centro-Oeste;

b) reportagens sobre turismo na América Central ou na América do Sul;

c) artigos científicos sobre epidemia do vírus Ebola e do vírus HIV fora do continente africano.

Resposta comentada

Respostas:

a) "artigo de revisão" E "política industrial" E ("região Sudeste" OU "região Centro-Oeste");

b) "reportagem" E turismo E ("América Central" OU "América do Sul");

c) "artigo científico" E epidemia E (Ebola OU HIV) NÃO (África OU continente africano).

2.7 ÁLGEBRA BOOLEANA E SISTEMAS DE RECUPERAÇÃO DA INFORMAÇÃO

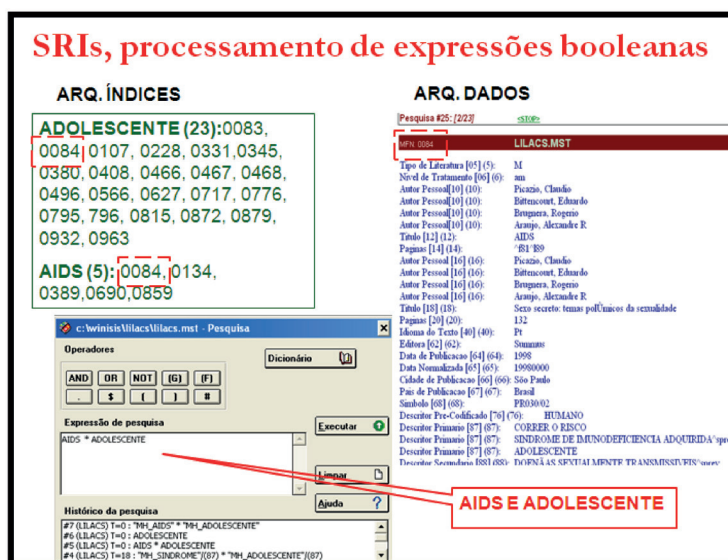
Os operadores booleanos são a base do processamento de consultas em sistemas de recuperação da informação computadorizados, porque são facilmente implementados como rotinas computacionais que realizam as operações booleanas entre conjuntos de índices de uma base de dados.



Na Figura 35, podemos ver esquematicamente três componentes de um SRI:

- a) a sua base de dados ou arquivo de dados (ARQ. DADOS);
- b) um dos seus arquivos de índices (ARQ. ÍNDICES);
- c) a sua interface de consulta (janela com tarja azul).

Figura 35 – Processamento de expressões booleanas em SRI fazendo uso de índices



Fonte: produção do próprio autor.

Nesse exemplo, é possível observar como o SRI pode responder à consulta "AIDS E ADOLESCENTE". Isso é feito porque as rotinas do SRI operam fazendo a intersecção dos índices de AIDS e de ADOLESCENTE; o resultado da operação E, ou intersecção dos índices AIDS e ADOLESCENTE, é o *registro 0084*, o único que pertence simultaneamente aos dois.

A álgebra booleana é uma linguagem de consulta que, de forma explícita ou não, nas interfaces, está presente em praticamente todos os SRIs, como veremos adiante.

Nos primeiros SRIs, o usuário tinha que entrar com uma expressão booleana completa. Veja, na Figura 36, uma expressão booleana submetida ao *Dialog*, um dos SRIs pioneiros, em operação desde a década de 1970 (<<http://support.dialog.com/searchaids/success/>>):

Figura 36 – Expressão booleana

S (ECONOMIA AND NORDESTE) AND AU=SILVA?

Fonte: produção do próprio autor.

Na linguagem de consulta do *Dialog*, "S" significa *search*, ou seja, *procure*, e "AU=" significa *procure autores que tenham SILVA no seu nome*.

Nos SRIs atuais, com suas interfaces *web*, não é necessário escrever uma expressão booleana completa para realizar uma busca. As interfaces

dos SRIs simplificam a maneira de um usuário submeter a eles expressões booleanas.

Na Figura 37, podemos ver a interface de consulta da base de dados LILACS, um SRI especializado em documentos sobre Ciências da Saúde, sobre o qual já falamos na Unidade 1, lembra? Pode-se observar que o usuário entra com os termos que quer buscar (AIDS e ADOLESCENTE) e pode especificar, na caixa de opções, à esquerda, com qual operador booleano – *and*, *or*, *and not* – deseja que os termos sejam relacionados.

Figura 37 – Operadores booleanos especificados no formulário de consulta de um SRI, no caso, a base de dados LILACS

The screenshot shows the LILACS search interface. At the top, there is a header with the logo 'bvs biblioteca virtual em saúde' and the text 'Pesquisa em bases de dados'. Below the header, there is a navigation bar with 'AJUDA', 'Base de dados: LILACS', and 'Formulário avançado'. The main search area is divided into two columns: 'Pesquisar' and 'no campo'. The 'Pesquisar' column has three rows: row 1 with 'AIDS', row 2 with 'and' and 'ADOLESCENTE', and row 3 with 'or' and 'and not'. The 'no campo' column has three rows: row 1 with 'Descritor de assunto', row 2 with 'Descritor de assunto', and row 3 with 'Palavras'. To the right of each row in the 'no campo' column is a small triangle icon and the word 'índice'. At the bottom of the search area, there are three buttons: 'CONFIG', 'LIMPAR', and 'PESQUISAR'.

Fonte: produção do próprio autor.

Algumas interfaces de busca de muitos SRIs são mais simples, como as que veremos a seguir.

Na Figura 38, podemos ver o formulário básico de pesquisa da base LILACS. Nele, o usuário só pode entrar com uma ou mais palavras e especificar os operadores booleanos E (todas as palavras) ou OU (qualquer palavra).

Figura 38 – Formulário de consulta da base LILACS com apenas as opções AND ou OR entre as palavras entradas

The screenshot shows the LILACS search interface. At the top, there is a header with the logo 'bvs biblioteca virtual em saúde' and the text 'Pesquisa em bases de dados'. Below the header, there is a navigation bar with 'AJUDA', 'Base de dados: LILACS', and 'Formulário livre'. The main search area has a text input field with the placeholder text 'Entre uma ou mais palavras'. Below the input field, there are two radio buttons: 'Todas as palavras (AND)' and 'Qualquer palavra (OR)'. At the bottom of the search area, there are two buttons: 'CONFIG' and 'PESQUISAR'.

Fonte: produção do próprio autor.

A Figura 39 mostra a interface de busca do catálogo da Library of Congress (LC). Nesse caso, se o usuário entrar com mais de uma palavra, o SRI sempre processará as palavras como se elas estivessem conectadas pelo operador booleano E; não há opções para os operadores OU e NÃO.

Library of Congress

Segundo a própria *Biblioteca do Congresso dos Estados Unidos da América*, ela é a maior biblioteca do mundo, com milhões de itens, incluindo livros, gravações, fotografias, mapas e manuscritos em suas coleções. A biblioteca ocupa três edifícios no *Capitólio: o Thomas Jefferson* (1897), o *John Adams* (1938) e o *James Madison Memorial Building* (1981).

Fonte: LIBRARY..., c2017.

Figura 39 – No catálogo de busca da *Library of Congress*, o formulário de consulta não oferece a opção de escolha de operadores booleanos



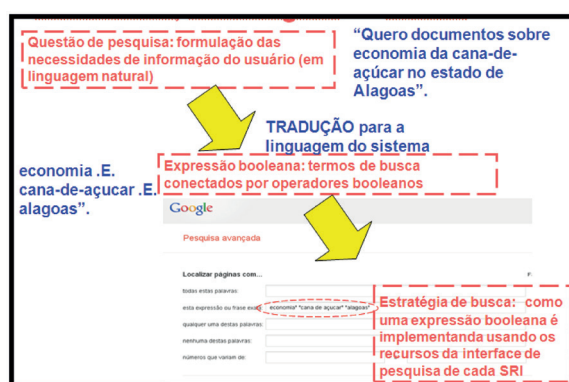
Fonte: captura de tela da página do catálogo de busca da *Library of Congress*.

Então, uma boa dica para se ter sucesso ao consultar qualquer SRI é usar um procedimento em três passos:

- tentar converter as demandas por informação em linguagem natural para uma expressão booleana;
- converter a expressão booleana numa estratégia de busca específica para a interface do SRI;
- submeter essa estratégia de busca ao SRI.

A Figura 40 resume as dicas anteriores, observe:

Figura 40 – Tradução de uma demanda por informação em linguagem natural para uma expressão booleana e para uma estratégia de busca em SRI



Fonte: produção do próprio autor.

Se é possível dizer que a álgebra booleana é a linguagem universal de pesquisa dos SRIs, seria interessante saber como traduzir sistematicamente para ela uma demanda por informação expressa em linguagem natural. Com isso, teríamos uma primeira aproximação para consultarmos determinado SRI, ou para submetermos a mesma demanda por informação a diferentes SRIs, um caso muito comum quando estamos avaliando qual, entre diversos sistemas, atende melhor a uma demanda por informações. Mas como fazer isso? Veja o passo a passo a seguir, no Quadro 1:

Quadro 1 – Como traduzir uma demanda por informação em linguagem natural para a linguagem da álgebra booleana

Como traduzir uma demanda por informação em linguagem natural para a linguagem da álgebra booleana
<ul style="list-style-type: none"> • Identifique os conceitos contidos na questão de pesquisa. • Desdobre os conceitos em um ou mais termos de busca. • Relacione os termos referentes a conceitos diferentes com o operador booleano E. • Relacione os termos referentes ao mesmo conceito com o operador booleano OU. • Exclua, com o operador booleano NÃO, aspectos não desejáveis. • Priorize os resultados dos termos de busca relacionados por operadores booleanos usando parênteses.

Fonte: produção do próprio autor.

A seguir, você encontrará alguns quadros e exemplos da aplicação dos princípios estudados até aqui, inclusive da análise conceitual (usando as categorias propostas pelo CRG), para auxiliar na tradução de demandas por informação expressas em linguagem natural para expressões em lógica booleana.

a) *Exemplo 1* – expressão de consulta: *cultura negra no Brasil*. Análise conceitual (Quadros 2 e 3):

Quadro 2 – Análise conceitual – CRG (Exemplo 1)

CATEGORIAS CRG	CONCEITOS
Coisas, substâncias, entidades	Cultura negra
Suas partes	
Sistemas de coisas	
Atributos de coisas	
Objetos da ação (paciente)	
Relações entre coisas, interações	
Operações sobre coisas	
Propriedades de atributos, relações e operações	
Lugar	Brasil
Tempo	

Fonte: produção do próprio autor.

Quadro 3 – Conceitos e termos (Exemplo 1)

CONCEITOS	TERMOS
Cultura negra	Cultura negra, cultura afro
Brasil	Brasil, brasileira

Fonte: produção do próprio autor.

Origami

Figura 41 – Origami



Fonte: Wikipédia.¹²

Técnica japonesa reconhecida como a arte de dobrar papel. O termo significa fazer dobras de papel, sem cortes nem colas, criando formas geométricas simples, mas podendo, também, gerar objetos ou seres de formatos complexos.

Berimbau

Figura 42 – Berimbau



Fonte: Wikipédia.¹²

Instrumento de origem africana que consiste em um arco de madeira retesado por uma corda de aço ou fio de arame. É ainda constituído de uma cabaça (caixa de ressonância), que fica presa à extremidade inferior da verga (arco de madeira) e modifica o som, funcionando como um ressoador.



Expressão booleana: *(cultura negra OU cultura afro) E (Brasil OU brasileira)*.

Observe que, nesse exemplo, foram utilizadas as categorias propostas pelo CRG como auxiliares para a identificação de conceitos. Isso não é obrigatório, mas pode, eventualmente, ajudar. “Cultura negra” foi identificado como um conceito único porque, se observarmos um origami ou um berimbau, saberemos que o primeiro é uma manifestação da cultura japonesa, enquanto o segundo, uma manifestação da cultura negra. O desdobramento do conceito de “Brasil” nos termos *Brasil* e *brasileira* ajuda que recuperemos um conjunto maior de documentos em que os conceitos estão expressos por termos diferentes.



Atenção

O uso de sinônimos só é válido em SRIs que não possuem *controle terminológico*, como os mecanismos de busca gerais da *web*. Em sistemas que utilizam vocabulário controlado, como a base de dados LILACS, por exemplo, que utiliza o vocabulário controlado DECS, devemos procurar pelo termo específico para cada conceito. Nesse caso, procure em: <http://decs.bvs.br/>.

b) *Exemplo 2* – expressão de consulta: *exercícios físicos na prevenção de doenças cardiovasculares*. Análise conceitual (Quadros 4 e 5):

Quadro 4 – Análise conceitual – CRG (Exemplo 2)

CATEGORIAS CRG	CONCEITOS
Coisas, substâncias, entidades	Exercícios físicos
Suas partes	
Sistemas de coisas	
Atributos de coisas	
Objetos da ação (paciente)	
Relações entre coisas, interações	Doenças cardiovasculares
Operações sobre coisas	
Propriedades de atributos, relações e operações	Prevenção
Lugar	
Tempo	

Fonte: produção do próprio autor.

¹⁰ Autor: *Andreas Bauer*. Disponível em: <https://en.wikipedia.org/wiki/Origami#/media/File:Origami-crane.jpg>. Acesso em: 30 jul. 2021.

¹¹ Autor: *Horst Nogajski*. Disponível em: https://en.wikipedia.org/wiki/Berimbau#/media/File:Hn_3berimbau.jpg. Acesso em: 30 jul. 2021.

Quadro 5 – Conceitos e termos (Exemplo 2)

CONCEITOS	TERMOS
Exercícios físicos	Exercícios físicos
Prevenção	Prevenção
Doenças cardiovasculares	Doenças cardiovasculares

Fonte: produção do próprio autor.

Expressão booleana: “exercícios físicos” E prevenção E “doenças cardiovasculares”.

Observe que os conceitos identificados estão entre aspas para deixar claro que são conceitos e não palavras isoladas.

- c) *Exemplo 3* – expressão de consulta: *locais de realização de provas do ENEM 2014 no Grande Rio*.

Neste exemplo, não vamos utilizar as categorias propostas pelo CRG para auxiliar. Como explicado anteriormente, essa etapa não é obrigatória para o desenvolvimento da expressão booleana.

Análise conceitual (Quadro 6):

Quadro 6 – Conceitos e termos (Exemplo 3)

CONCEITOS	TERMOS
ENEM	ENEM, Exame Nacional do Ensino Médio
2014	2014
Prova	Prova
Grande Rio	Grande Rio, Rio de Janeiro, Caxias, Nova Iguaçu, São João de Meriti, Belford Roxo, Niterói, São Gonçalo

Fonte: produção do próprio autor.

Expressão booleana: (ENEM OU “Exame Nacional do Ensino Médio”) E 2014 E prova E (“Grande Rio” OU “Rio de Janeiro” OU Caxias OU “Nova Iguaçu”...).

Atente que “Grande Rio” é um coletivo. Para que possamos recuperar o máximo de informações possível, por exemplo, uma notícia no *Google* sobre “locais de realização do ENEM 2014 em São João de Meriti”, devemos *desdobrar um coletivo* em seus componentes. Se não o fizermos, utilizando somente “Grande Rio”, essa notícia específica não será recuperada, assim como não serão recuperadas tantas outras que poderiam ser relevantes.

- d) *Exemplo 4* – expressão de consulta: *desmatamento na Amazônia nos últimos três anos*. Análise conceitual (Quadro 7):



Quadro 7 – Conceitos e termos (Exemplo 4)

CONCEITOS	TERMOS
Desmatamento	Desmatamento
Amazônia	Amazônia
Últimos três anos	Últimos três anos, 2012, 2013, 2014

Fonte: produção do próprio autor.

Expressão booleana: *desmatamento E Amazônia E (“últimos três anos” OU 2012 OU 2013 OU 2014)*.

Imagine que você está procurando notícias sobre desmatamento na Amazônia nos últimos três anos. Repare que o conceito “últimos três anos” é o OU de notícias sobre o desmatamento da Amazônia em 2012, em 2013 e em 2014. Também podem aparecer notícias que falem da quantificação do desmatamento na Amazônia exatamente nos três últimos anos. A expressão de busca sugerida recupera todas essas notícias, sejam elas específicas, de qualquer um dos três últimos anos, ou dos três últimos anos em conjunto.



2.7.1 Atividade

Se for possível, tente fazer esta atividade a partir de um dispositivo com acesso à internet, usando o *Google*.

Formule as expressões booleanas equivalentes às expressões de consulta a seguir. Para facilitar sua análise, identifique, nos quadros disponibilizados, os conceitos e termos existentes em cada uma:

- técnicas de agricultura orgânica para cultivo de hortaliças em cobertura de prédios;

Quadro 8 – Conceitos e termos

CONCEITOS	TERMOS
agricultura orgânica	agricultura orgânica
hortaliça	hortaliça
cobertura de prédio	cobertura de prédio

Fonte: produção do próprio autor.

- prefeituras de cidades do estado de Santa Catarina governadas pelo Partido dos Trabalhadores (PT), exceto a capital;

Quadro 9 – Conceitos e termos

CONCEITOS	TERMOS
prefeitura	prefeitura
Santa Catarina	Santa Catarina, SC
Partido do Trabalhadores	Partido do Trabalhadores, PT
Florianópolis	Florianópolis, capital

Fonte: produção do próprio autor.

- c) encontre usuários que sejam engenheiros, exceto os engenheiros mecânicos, e que tenham residência nos bairros de Icaraí ou São Francisco, na cidade de Niterói.

Quadro 10 – Conceitos e termos

CONCEITOS	TERMOS
engenheiro	engenheiro
engenheiro mecânico	engenheiro mecânico
residente	residente, domiciliado, endereço
Icaraí	Icaraí
São Francisco	São Francisco
Niterói	Niterói

Fonte: produção do próprio autor.

Resposta comentada

Respostas:

- “agricultura orgânica” E hortaliça E “cobertura de prédio”;
- prefeitura E (“Santa Catarina” OU SC) E (Partido dos Trabalhadores OU PT) NÃO (Florianópolis OU capital);
- engenheiro E “engenheiro mecânico” E (residente OU domiciliado OU endereço) E (Icaraí OU “São Francisco”) E Niterói.



2.8 CONCLUSÃO

Embora aparentemente simples e amplamente utilizados no dia a dia das pessoas, os SRIs são sistemas complexos. O que enxergamos deles é somente sua interface de consulta, uma simples caixa de busca onde entramos com palavras-chave que expressam nossas demandas por informação. Nunca devemos consultar um SRI imaginando que ele é uma pessoa que compreende nossa linguagem, embora muitos SRIs pareçam fazê-lo.

A análise conceitual de demandas por informação, expressas pelos usuários em linguagem, é um instrumento fundamental para que tenhamos sucesso em nossas interações com diversos tipos de SRIs. Ela vai nos instrumentalizar para formularmos expressões em lógica booleana, a linguagem que é compreensível pela maioria dos SRIs, com mais precisão, aumentando a chance de que nossas consultas a esses sistemas sejam mais proveitosas.

RESUMO

SRIs são intermediários entre demandas por informação e uma coleção de registros. Para que sua utilização seja a mais proveitosa possível, é necessário que consigamos especificar, com o máximo de clareza e precisão, nossas demandas por informação.

Essas demandas são pensadas e formuladas em linguagem natural. Acontece que os SRIs não a compreendem, embora muitos deles possam parecer fazê-lo. Os registros informacionais que recuperamos dos SRIs têm propriedades específicas, como autor, título e, principalmente, tema ou assunto, que precisamos especificar claramente para que os sistemas compreendam. Essa especificação é feita pela descrição das propriedades dos itens de informação, por meio de *unidades de pensamento*, *conceitos* ou *ideias*, que são expressas através de *termos* linguísticos. É preciso, então, que conheçamos os instrumentos teóricos para analisarmos uma demanda por informações em linguagem natural e saibamos identificar aí os conceitos/termos.

Quando a especificação da informação que desejamos consta de várias propriedades, expressas por diversos conceitos/termos, usamos os operadores booleanos E, OU e NÃO para expressar as diferentes relações entre os termos que formam a especificação de nossa consulta.

REFERÊNCIAS

CHIERCHIA, Gennaro. **Semântica**. Campinas: UNICAMP, 2003.

GEORGE Boole. **George Boole 200**, Cork, c2017. Disponível em: <<http://georgeboole.com/>>. Acesso em: 21 jan. 2015.

LIBRARY of Congress. **Library of Congress**, Washington, c2017. Disponível em: <<https://www.loc.gov/>>. Acesso em: 21 nov. 2015.

RANGANATHAN, S. R. **Prolegomena to a library classification**. 3rd ed. Bombay: Asia Publishing House, 1967.



Sugestão de Leitura

CLEVERDON, C. W. **The effect of variations in relevance assessments in comparative experimental tests of index languages**. Cranfield: Cranfield Institute of Technology, 1970. (Cranfield Library Report, 3).

CLEVERDON, C. W.; MILLS, J.; KEEN, E. M. **Factors determining the performance of indexing systems**. Cranfield: Aslib Cranfield Research Project (College of Aeronautics), 1966. v. 1-2.

DAHLBERG, Ingetraut. Teoria do Conceito. **Ciência da Informação**, Brasília, v. 7, n. 2, p. 101-107, 1978. Disponível em: <<http://revista.ibict.br/cienciadainformacao/index.php/ciinf/article/viewFile/1680/1286>>. Acesso em: 9 nov. 2012.

FILLMORE, C. J. **The case for case**. [S.l.]: Universals in Linguistic Theory, 1968. p. 1-88.

INFORMATION retrieval. **Wikipedia**, [S.l.], 2017. Disponível em: <https://en.wikipedia.org/wiki/Information_retrieval>. Acesso em: 13 abr. 2021.

RELEVANCE: information retrieval. **Wikipedia**, [S.l.], 2017. Disponível em: <[https://en.wikipedia.org/wiki/Relevance_\(information_retrieval\)](https://en.wikipedia.org/wiki/Relevance_(information_retrieval))>. Acesso em: 13 abr. 2021.

SARACEVIC, Tefko. Relevance: a review of the literature and a framework for thinking on the notion in information science: part II: nature and manifestations of relevance. **Journal of the American Society for Information Science and Technology**, [S.l.], Nov. 2007. Disponível em: <<https://comminfo.rutgers.edu/~tefko/Saracevic%20relevance%20pt%20II%20JASIST%20%2707.pdf>>. Acesso em: 7 set. 2012.





SMIT, Johanna (Coord.). **Análise documentária**: a análise da síntese. Brasília: IBICT, 1987.

SMITHSON, Steve. Information retrieval evaluation in practice: a case study approach. **Information Processing and Management**: An International Journal, [S.l.], v. 30, n. 2, p. 205–221, Mar./Apr. 1994. Disponível em: <<http://dl.acm.org/citation.cfm?id=175462>>. Acesso em: 29 jun. 2015.
