

## LÓGICA CLÁSSICA E LÓGICAS NÃO-CLÁSSICAS NOS MODELOS DE RECUPERAÇÃO DA INFORMAÇÃO

José Claudio Morelli Matos

 <http://lattes.cnpq.br/5814859475637460> –  <https://orcid.org/0000-0002-7763-4971>  
[doutortodd@gmail.com](mailto:doutortodd@gmail.com)

Universidade do Estado de Santa Catarina (UDESC)  
Florianópolis, SC, Brasil

### RESUMO

A utilização da lógica nos modelos de recuperação da informação (RI) é examinada, por meio de levantamento bibliográfico da produção no campo e análise crítica de trabalhos relevantes. O levantamento revela um crescimento do volume de pesquisas acerca de modelos não-clássicos que incorporam a incerteza, especialmente a lógica difusa (*fuzzy logic*) nos sistemas de RI. A evolução de sistemas baseados em lógicas não-clássicas, emergidas a partir da lógica clássica/booleana é discutida. Conclui com uma tentativa de responder a críticas contra o formalismo lógico e reconhece a evolução técnica e conceitual representada pela introdução de modelos lógicos não-clássicos no campo da RI.

**Palavras-chave:** Recuperação da informação. Lógica. Lógica difusa.

### CLASSICAL LOGIC AND NON-CLASSICAL LOGICS IN INFORMATION RETRIEVAL MODELS

#### ABSTRACT

The use of logic in information retrieval (IR) models is examined through a bibliographic survey of the production in the field and critical analysis of relevant works. The survey reveals a growing volume of research on non-classical models that incorporate uncertainty, especially fuzzy logic in IR systems. The evolution of systems based on non-classical logics emerged from the classical/Boolean logic is discussed. It concludes with an attempt to respond to criticisms against logical formalism and recognizes the technical and conceptual evolution represented by the introduction of non-classical logical models in the field of IR.

**Keywords:** Information retrieval. Logic. Fuzzy logic.

### LÓGICA CLÁSICA Y LÓGICAS NO CLÁSICAS EN MODELOS DE RECUPERACIÓN DE LA INFORMACIÓN

#### RESUMEN

El uso de la lógica en los modelos de recuperación de la información (RI) se examina a través de un estudio bibliográfico de la producción en el campo y el análisis crítico de trabajos relevantes. La encuesta revela un volumen creciente de investigación sobre modelos no clásicos que incorporan incertidumbre, especialmente lógica difusa en sistemas de RI. Se discute la evolución de los sistemas basados en lógicas no clásicas a partir de la lógica clásica/booleana. Concluye con un intento de responder a las críticas contra el formalismo lógico y reconoce la evolución técnica y conceptual representada por la introducción de modelos lógicos no clásicos en el campo de las RI.

**Palabras clave:** Recuperación de la información. Lógica. Lógica difusa.

DOI <http://dx.doi.org/10.1590/1981-5344/48385>

Recebido em: 13/10/2023.

Aceito em: 08/05/2024.



## 1 INTRODUÇÃO

Este estudo pretende discutir a participação de diferentes sistemas de lógica na construção e utilização de Sistemas de Recuperação da Informação (SRI). A Recuperação da Informação (RI) é considerada uma área de estudos integrante do campo da Ciência da Informação (CI). Trata-se de um ambiente de pesquisas cuja principal atividade consiste em projetar, construir e avaliar sistemas automáticos para recuperação de documentos em coleções e bases de dados. Possui, por isso mesmo, uma forte ligação com as ciências da computação. É possível afirmar que, como área de conhecimento e pesquisas, a RI seja um território de fronteira entre a CI e a ciência da computação.

Ao se realizar uma busca por artigos sobre RI, em bases de dados científicos como a *Web of Science* (WOS), é possível perceber pelos resultados, que são muito comuns os artigos que tratam de implantação e teste de SRI. Reflexões sobre conceitos teóricos e crítica de pontos de vista são menos comuns. Explicações conceituais e reflexões críticas no campo da RI são encontradas geralmente em livros-texto, em que tais ideias vêm a ser transmitidas aos aprendizes deste campo de conhecimento. Na tentativa de empreender uma investigação conceitual de caráter mais reflexivo sobre a RI, a presente discussão trata dos modelos lógicos de recuperação. A maioria desses modelos tem sua origem na álgebra booleana, mas o escopo deste estudo estende-se a toda a lógica clássica de inspiração matemática. Em sentido mais amplo, inclui também os sistemas conhecidos como lógica difusa (*fuzzy logic*) e as demais lógicas não-clássicas.

O argumento, portanto, toma como ponto de partida a profunda fundamentação dos SRI na álgebra booleana. É discutida uma concepção instrumental da lógica, a partir de uma revisão de alguns trabalhos teóricos neste campo. A relação destes sistemas lógicos com os fundamentos da RI é estabelecida através da revisão da produção relevante neste assunto. A problemática de que trata este estudo poderia ser formulada nos termos das seguintes questões:

- De que forma é possível conceber a relação da lógica com a recuperação da informação?
- Como as diferenças entre a lógica clássica e as lógicas não-clássicas podem ser consideradas no contexto desta relação?

## 2 PROCEDIMENTOS METODOLÓGICOS

A fim de responder às questões formuladas, o procedimento escolhido para a amostragem foi a pesquisa bibliográfica. Foi elaborada uma estratégia de busca para recuperar artigos científicos, em que a relação entre lógica e RI ficasse evidenciada. Esta estratégia recorreu à base de dados científicos WOS, para a pesquisa com os termos “lógica” (*logic*) e “recuperação da informação” (*information retrieval*), conectados pelo operador booleano de conjunção (AND). Em seguida, foram aplicadas as seguintes operações de refinamento:

1. O intervalo de tempo foi demarcado entre 1981 e 2020, compreendendo 40 anos de produção.
2. As áreas do conhecimento foram demarcadas compreendendo a Biblioteconomia e Ciência da Informação (BCI) e campos da computação científica e aplicada. Outros campos do conhecimento foram excluídos do universo da pesquisa.
3. Foram recuperados somente documentos na forma de artigos em periódicos, descartando demais tipos de produção.

A busca realizada conforme esta estratégia recuperou 293 artigos. Este resultado foi exportado para uma planilha do programa *Excell*, na ordem do mais recente para o mais antigo, incluindo dados como título, autoria, resumo, ano e veículo de publicação. Nesta lista de artigos, foram realizadas as seguintes operações de codificação e análise:

1. Foram listados todos os artigos em que o termo “lógica” (*logic*) aparece no título. Esta lista constitui a categoria “lógica” em sentido geral, que inclui a álgebra booleana.
2. Foram listados todos os artigos em que o termo “difusa” (*fuzzy*) aparece no título, relacionado com o termo “lógica” (*logic*). Esta lista constitui a categoria “lógica difusa” (*fuzzy logic*).
3. Em artigos cujo título não oferecia informação suficiente ou clara, foi feita a análise do resumo para decidir pela inclusão ou exclusão da lista.
4. Foi registrado o ano de publicação dos artigos em cada uma das listas e elaborado um quadro com o número de artigos de cada categoria por ano e por década (Quadro 1). O cumprimento

desta etapa, por si só, já produziu alguns resultados relevantes para o assunto do estudo. A etapa seguinte completou o procedimento metodológico.

5. Dos artigos listados em ambas as categorias, através da análise dos resumos, foram selecionados 20 artigos a fim de serem analisados na íntegra, para formar a base argumentativa do presente estudo. O critério de escolha nesta etapa foi a identificação da presença de noções claras e bem desenvolvidas sobre a participação de modelos lógicos no campo da RI.

**Quadro 1** – quantitativo de artigos por ano e somatório por década

Ano	Lógica formal	Total da década	Lógicas não-clássicas – fuzzy	Total da década
2020	0	7	4	22
2019	1		4	
2018	1		3	
2017	2		4	
2016	0		2	
2015	2		3	
2014	0		1	
2013	1		1	
2012	0		0	
2011	0		0	
2010	0	8	2	19
2009	0		4	
2008	0		2	
2007	1		1	
2006	2		2	
2005	0		4	
2004	1		0	
2003	1		0	
2002	0		2	
2001	3		2	
2000	2	14	1	6
1999	4		0	
1998	3		1	
1997	1		1	
1996	0		2	
1995	2		1	
1994	1		0	
1993	0		0	
1992	1		0	
1991	0		0	
1990	0	3	0	3
1989	1		1	
1988	0		0	
1987	1		0	
1986	0		1	
1985	0		0	
1984	0		0	
1983	1		0	
1982	0		0	

1981	0		1	
------	---	--	---	--

Fonte: elaborado pelo autor (2020).

Observando o Quadro 1, pode-se perceber uma relação invertida de crescimento do volume de publicações entre os artigos da primeira categoria (lógica) e da segunda (lógica difusa). Descartando-se a primeira década – os anos 1980 – tem-se um número decrescente de artigos na amostra, na categoria “lógica” nas três décadas seguintes e um número crescente de artigos na categoria “lógica difusa”. Este crescimento de interesse por modelos de RI baseados na lógica difusa será analisado nas seções seguintes do estudo, e suas causas e implicações receberão tratamento crítico, com base na bibliografia recuperada na etapa de amostragem.

### 3 O CAMPO DE ESTUDOS DA RECUPERAÇÃO DA INFORMAÇÃO

Antes de discutir os fundamentos da RI como área de pesquisa especializada, convém compreender melhor o seu objeto de estudo. Na obra *An Introduction to Information Retrieval*, os autores oferecem a seguinte definição:

Recuperação da informação (RI) é encontrar material (usualmente documentos) de uma natureza não-estruturada (usualmente texto) que satisfaça uma necessidade de informação, dentro de grandes coleções (usualmente armazenadas em computadores) (Manning; Raghavan; Schütze, 2009, p. 01, tradução nossa).

Perceba-se que usam o advérbio “usualmente”, querendo dizer que documentos em forma de textos armazenados em computadores são o objeto mais comum dos SRI, embora objetos de outra natureza, arquivos de imagem, som e vídeo, também possam ser armazenados de forma que sua recuperação também envolva algum SRI.

Manning, Raghavan e Schütze (2009) acrescentam que, com as mudanças causadas pela tecnologia digital, esta atividade, que antes atendia a algumas poucas pessoas, atualmente se tornou parte da realidade de grandes parcelas da população. Afirmam eles que:

Agora o mundo mudou e centenas de milhões de pessoas se engajam na recuperação de informação todos os dias quando usam um motor de busca na web ou verificam seu e-mail (Manning; Raghavan; Schütze, 2009, p. 01, tradução nossa).

Programas conhecidos como “motores de busca” (*search engines*) tornaram-se habitantes de um novo e complexo ecossistema informacional,

realizando a ligação entre os usuários e as coleções de documentos. O desenvolvimento de sistemas rápidos e fáceis de manusear fez com que os motores de busca se tornassem os principais modos de recuperar informação usados atualmente.

A RI é uma das áreas, se não for a principal, que mais aplica princípios e sistemas lógicos, em todo o campo da CI. Faz uso, em sua base, de modelos binários de lógica, como a álgebra booleana, lógica clássica em geral e teoria de conjuntos. Contudo, a diversidade de aplicações exige frequentemente o emprego de extensões que se baseiam em sistemas de lógica não-clássica, em que regras e princípios da lógica clássica são modificados ou substituídos, de forma a aumentar sua aplicabilidade na RI. Sobre o desenvolvimento da lógica e sua diversificação recente nas aplicações computacionais, vide Aucher (2017), capítulo 1. Acerca dos sistemas lógicos, conforme D'Ottaviano e Feitosa (2003, p. 22): “Existem vários outros tipos de lógicas não-clássicas como, por exemplo, as *lógicas difusas (lógicas fuzzy)*”. No processo evolutivo dos modelos e sistemas, a incorporação de modelos lógicos não-clássicos é sinal de avanço teórico e técnico do campo da RI.

#### **4 LÓGICA E MODELOS NA RI**

Embora a lógica como disciplina tenha origem na tradição filosófica que remonta a Aristóteles, a participação da lógica na RI quase nada tem a ver com a filosofia. Está muito mais assentada nos modelos de prova e representação formal de informações, que caracterizam a lógica matemática desde o final do século XIX. Esta modelagem formal, inspirada nas matemáticas, tem como seu marco fundante o trabalho de George Boole, com sua álgebra lógica, e não é por acaso que os modelos mais básicos de RI se fundamentam neste sistema de lógica, conhecido por sua simplicidade e exatidão. Mas o avanço da tecnologia digital e das pesquisas - teóricas e aplicadas - no campo da RI acabaram estabelecendo novas exigências para os sistemas, assim como para sua representação teórica em modelos. Sobre a relação entre lógica e modelos, no caso específico da CI, vide Mahr (2009). Desdobramentos mais recentes da lógica passaram a ser considerados, para permitir a construção de SRI mais eficientes e lidar com novas e mais complexas necessidades dos usuários.

Foi a partir do artigo *A non-classical logic for information retrieval* (Van Rijsbergen, 1986), que a RI passa a contar com uma teoria consistente sobre a utilização de modelos lógicos. O autor afirma que, na modelagem lógica clássica, de inspiração matemática, a exemplo dos modelos booleanos básicos, “documentos são representados por conjuntos de termos do índice” (Van Rijsbergen, 1986, p. 06, tradução nossa). A fórmula conjuntiva dos termos do índice, que o documento contém em seu texto, é a representação “lógica” deste documento. A operação de recuperação, nos moldes básicos da busca booleana, pode ser formalizada ou modelada com os conectivos da álgebra booleana. Observa-se também o uso da linguagem da teoria clássica de conjuntos.

A necessidade de modelos é um pressuposto profundamente assumido por todos os principais autores que discutem aspectos lógicos da RI. Há consenso em torno da noção de que as linguagens lógicas fornecem rigor, consistência e capacidade preditiva aos modelos. Van Rijsbergen (1986, p. 04, tradução nossa) afirma que a modelagem em RI necessita atingir o objetivo de “especificar, de uma vez por todas, qual a relação entre um documento e uma consulta”. Essa questão central é reiterada por autores como Chiaramella e Chevallet (1992), em seu artigo *About Retrieval Models and Logic*, os autores acrescentam uma detalhada explicação da função e das características dos modelos em RI. Segundo estes, “os modelos de recuperação constituem o mais poderoso meio de comparar as potencialidades de vários sistemas de recuperação” (Chiaramella; Chevallet, 1992, p. 233, tradução nossa). A função básica da lógica na RI é fornecer modelos. Os modelos permitem construção de sistemas, mas também sua análise formal.

Lalmas (1998, p. 19, tradução nossa) afirma que “alguns modelos lógicos são capazes de representar, em uma estrutura uniforme, vários aspectos de sistemas de RI, assim como a semântica da informação”. A precisão e exatidão semântica das teorias formais da lógica fornece um instrumento para a construção e análise de modelos de RI, assim como para a previsão teórica de seu desempenho. Esta é uma vantagem enfatizada pelos autores que tratam do assunto. A semântica rigorosa das linguagens lógicas, como já se pode perceber mesmo na lógica booleana/proposicional, permite uma

representação das operações de recuperação, de modo a decidir que características do modelo devem ser mantidas e quais devem ser modificadas.

Sebastiani (1998) especifica três operações de modelagem: descritiva, prescritiva e preditiva. Modelos descritivos formalizam os aspectos principais do sistema, mesmo de forma simplificada. Os modelos do tipo prescritivo procuram influenciar a futura construção de modelos. A preditiva, específica da RI, constitui formalização, ou experimentação abstrata, para analisar e testar os sistemas modelados. Segundo o autor, "um modelo lógico deve ser construído incrementalmente" (Sebastiani, 1998, p. 03, tradução nossa). Se for assim – incrementalmente – que os modelos são tratados pela IR, é mais importante preservar a adequação formal do que a compreensividade (cobertura completa de aspectos) do modelo. Essa é a postura básica segundo a qual modelos lógicos devem evoluir, a fim de que as simplificações venham (incrementalmente) a dar lugar a funcionalidades cada vez mais complexas, na base do modelo lógico adotado. Este incremento podendo levar à mudança deste modelo para outros, que representem melhor a função entre a relevância e a necessidade dos usuários.

Segundo Sebastiani (1998), no modelo proposto por Van Rijsbergen (1986),

[...] o papel da probabilidade é chave, já que fórmulas lógicas são representações inerentemente imperfeitas de documentos e necessidades informacionais; a relevância de um documento para uma necessidade informacional, assim, pode ser estabelecida apenas sobre um grau limitado de certeza (Sebastiani, 1998, p. 04, tradução nossa).

O que o autor está dizendo é que o modelo lógico é um campo de experimentação mais interessante dos sistemas, para compreender a razão pela qual certas técnicas funcionam melhor que outras. Ele dá indicação, ao comparar a RI com a Inteligência Artificial (IA), de que a lógica aplicada é um método de avaliação de sistemas computacionais em geral. Sebastiani (1998, p. 03, tradução nossa) justifica esta generalização chamando atenção para a possibilidade de considerar a "lógica como uma linguagem comum", entre diferentes subcampos e comunidades de pesquisadores. Talvez valesse a pena acrescentar que os modelos lógicos fornecem uma linguagem comum entre os pesquisadores e as máquinas que executam os sistemas.

Voltando ao trabalho de Van Rijsbergen (1986), seu modelo parte da fórmula que representa a relação entre o documento e a consulta. Dado um documento  $d$ , presente em uma coleção, e uma consulta  $q$ , feita por um usuário em um SRI, a relação formal entre eles é representada por uma versão da operação de *implicação*. A fórmula pode ser escrita assim:

$$d \rightarrow q$$

Segundo o autor: "Esta relação pode ser descrita em termos da probabilidade da implicação material" (Van Rijsbergen, 1986, p. 04, tradução nossa). Esta é a originalidade da descoberta de Van Rijsbergen (1986): a noção de implicação da lógica clássica atende a finalidades relacionadas com a fundamentação e representação de noções da matemática. Na construção de modelos de RI esta noção precisa ser reformulada a fim de evitar inconsistências entre o documento e a consulta.

Usando a noção clássica de implicação material, não é possível operar com limites de incerteza e probabilidade, como é o caso da relação entre o documento e a consulta formulada pelo usuário para manifestar sua *necessidade* de informação. Por isso, se justifica a necessidade de uma "nova estrutura para a Recuperação da Informação baseado na lógica não-clássica" (Van Rijsbergen, 1986, p. 12, tradução nossa). Em 1986, antes do estabelecimento de uma produção robusta sobre a lógica na RI, Van Rijsbergen (1986) já discutia qual o modelo lógico que melhor representaria probabilidade na fórmula em que ocorre o conectivo de implicação. Este autor tinha uma postura crítica em relação à implicação material adotada na lógica clássica de primeira ordem, e sugeria uma abordagem não-clássica para este aspecto do modelo: o tipo de implicação existente entre o documento e a necessidade de informação que deu origem à busca.

A necessidade de representar formalmente a imprecisão ou incerteza, assim como formalizar graus de relevância para os documentos recuperados, estimula a elaboração de modelos de RI formulados e linguagens lógicas alternativas à lógica clássica. Lalmas (1998, p. 24, tradução nossa) afirma sobre isso que "a implicação  $d \rightarrow q$  tal como é definida na lógica clássica, não representa adequadamente a relevância de um documento para uma

consulta". No caso da RI, o principal modelo clássico é o modelo booleano. Sebastiani (1999) comenta justamente que o:

[...] modelo booleano tradicional da RI (em que a linguagem lógica para a representação de documentos é a das conjunções booleanas de letras proposicionais) é um modelo no qual o conhecimento total é implicitamente assumido" (Sebastiani, 1999, p. 413, tradução nossa).

A álgebra booleana, por seu aspecto proposicional, é um tipo de modelo de "assunção de conhecimento total" (Sebastiani, 1999, p. 413, tradução nossa), ou seja, tudo o que pode ser representado por essa linguagem no domínio do assunto em questão é assumido como conhecido. A sua limitação, neste aspecto, é a dificuldade de representar a incerteza, na relação entre a necessidade do usuário e o documento. Por esse motivo, lógicas que assumem conhecimento parcial, ao representar formalmente a incerteza (ou até contradição) são adotadas em modelos mais complexos de RI, a fim de permitir mais abrangência ou cobertura.

Recentemente foi publicado um trabalho de revisão sobre o papel da lógica na RI (Abdulahhad *et al.*, 2019). Neste artigo, os autores reconstroem a argumentação em favor da modelagem de SRI com uso das diversas linguagens lógicas, numa concepção pluralista que é típica do desenvolvimento da lógica no tocante às suas aplicações para a computação. O artigo afirma que:

[...] as lógicas formais são ferramentas poderosas para a representação do conhecimento, a integração do conhecimento nos processos de RI e para representar a natureza inferencial da decisão de recuperação (Abdulahhad *et al.*, 2019, p. 18, tradução nossa).

O conceito de inferência, que está no cerne da própria ideia de raciocínio e, portanto, da própria lógica, é tomado como fundamento para a representação da relação entre o documento e a consulta. O que o SRI faz, quando recupera um conjunto de documentos a partir da consulta do usuário, é uma inferência que manifesta certo grau de incerteza, que a lógica precisa poder representar formalmente.

Segundo os resultados do levantamento bibliográfico deste estudo, os modelos lógicos na RI seguem mais frequentemente algumas linguagens não-clássicas, reconhecidas pela sua potencialidade para representar a relação entre documento e consulta ( $d \rightarrow q$ ). Tais linguagens operam apreendendo o aspecto gradativo da relevância e a incerteza decorrente da ambiguidade da

linguagem natural, empregada na maioria dos documentos. Representantes dessas linguagens, que foram identificadas na amostra pesquisada são: a *default logic* (lógica de padrões), a *description logic* (lógica descritiva) e, como principal lógica não-clássica, com uma produção bastante robusta de pesquisas, a *fuzzy logic* (lógica difusa). Em seguida se passa a uma breve apresentação desses sistemas, na forma como são utilizados na RI.

## 5 AS LÓGICAS NÃO-CLÁSSICAS NA RI

Um dos sistemas lógicos que têm sido tratados na produção baseada na relação entre lógica e RI, conforme a pesquisa bibliográfica realizada neste estudo, é a lógica dos padrões (*default logic*, traduzida em alguns documentos em português para “lógica *default*”). É um sistema que complementa as operações da lógica proposicional com uma regra de inferência que permite representar graus de incerteza e relevância. Foi desenvolvida por Reiter (1980) e tem sido alvo do interesse de estudiosos da RI especialmente nas últimas duas décadas, conforme indicam os resultados do levantamento bibliográfico.

Hunter (2001), explica as linhas gerais desse sistema nos seguintes termos:

As regras de inferência são aquelas da lógica clássica mais um mecanismo especial para lidar com regras padrão: Basicamente, se  $\alpha$  é inferido, e  $\neg\beta$  não pode ser inferido, então inferir  $\gamma$ . Por isso,  $\alpha$  é chamado a pré-condição,  $\beta$  é chamado a justificação, e  $\gamma$  é chamado o conseqüente (Hunter, 2001, p. 66, tradução nossa).

Recorde-se que as letras  $\alpha$ ,  $\beta$  e  $\gamma$  correspondem a fórmulas bem formadas na linguagem da lógica proposicional clássica. Este esquema de regra padrão é representado assim:

$\alpha: \beta / \gamma$

O principal interesse deste sistema para a construção e análise de modelos de RI é poder representar as consultas como inferências, em que está envolvida certa incerteza, inerente à relação de relevância entre documentos e consultas. Antoniou (1999), no artigo *A Tutorial on Default Logics* explica o funcionamento da lógica dos padrões como a representação de uma linha de inferências feita por um sistema inteligente, baseada em um conhecimento que não é completo.

Nesses casos, segundo o autor, “o sistema tem que fazer algumas conjecturas plausíveis, que no caso do raciocínio por padrões, são baseadas em regras gerais, chamadas padrões” (Antoniou, 1999, p. 337, tradução nossa). O que traduz-se aqui por “regras gerais” é uma expressão do inglês: “*rules of thumb*”, poderia ser traduzida por “via de regra”, ou seja, uma regra baseada em uma inferência plausível, mas que não é fundamentada em um conhecimento completo. Conforme Hunter (2001, p. 67, tradução nossa): “A teoria dos padrões, então, aumenta estas inferências clássicas por meio das inferências padrão deriváveis, usando as regras padrão”.

Outro sistema lógico presente nos estudos relacionando RI e lógica é a lógica descritiva (*description logics*). Segundo Baader *et al.* (2003, p. 05, tradução nossa), é possível considerar a “lógica descritiva (LD) como um formalismo para representar conhecimento”. Seu formalismo complementa a lógica clássica de primeira ordem, introduzindo operadores para a representação de relações entre conceitos. Kaibo *et al.* (2015) descrevem seu funcionamento, num artigo em que discutem a capacidade deste modelo para representar informação em bases de dados. Afirmam os autores:

Os elementos básicos das lógicas descritivas são conceitos e regras, que denotam classes e relações binárias respectivamente. Construtores podem ser aplicados para formar expressões de conceitos. Os operadores de conjuntos usados são complemento ( $\neg$ ), união ( $\cup$ ) e intersecção ( $\cap$ ) que são denotados como negação, disjunção e conjunção” (Kaibo *et al.*, 2015, p. 282, tradução nossa).

Perceba-se que a base para a sistematização da lógica descritiva, mais uma vez, é a lógica clássica de primeira ordem, considerando a teoria clássica dos conjuntos como uma parte sua.

De todos os sistemas lógicos que vêm sendo identificados nas pesquisas em RI, contudo, são as espécies de *fuzzy logic* (lógica difusa, traduzida também como “lógica nebulosa”) que representam a maioria das ocorrências, especialmente nos anos mais recentes. Expressões frequentes nos artigos da amostra pesquisada incluem também *fuzzy sets* (conjuntos difusos) e *fuzzy rules* (regras difusas). A intenção subjacente a estes sistemas é satisfazer a necessidade de representar *graus* de pertencimento de determinado elemento a um conjunto. Krause (2003) comenta sobre isso no contexto das lógicas não clássicas, quando reconhece que em alguns tipos de conjuntos, um elemento pode estar “mais para dentro” do conjunto do que outro. Isso pode ser

explicado nos termos de que, na RI, um documento pode representar melhor uma consulta do que outro. Esta diferença, que a RI chama de relevância, pode ser analisada por uma teoria formal que reconheça a nebulosidade ou o aspecto difuso (*fuzzyness*) de resultados de inferências.

Bouidghaghen *et al.* (2009), refletem acerca da característica difusa da semântica dos textos em linguagem natural, a fim de introduzir a discussão sobre os sistemas de lógica difusa para o tratamento de textos em sistemas de RI. Afirmam:

Textos são difusos em muitos aspectos, mesmo quando seus autores não tiveram a intenção de ser vagos e antes tentaram ser acurados nas suas afirmações. Isto parece dever-se à própria natureza das linguagens naturais, que ecoam o modo como humanos percebem o mundo (Bouidghaghen *et al.*, 2009, p. 82, tradução nossa).

Se esta é uma qualidade dos textos e do pensamento humano, a recuperação de documentos em linguagem natural precisa se orientar por sistemas em que o aspecto difuso seja representado, sem perda de conhecimento. Os autores complementam afirmando que:

[...] conjuntos difusos proporcionam representações úteis de categorias associadas com propriedades graduais como 'largo', 'alto', 'jovem' ou 'barato', possivelmente modulados por demarcações linguísticas (Bouidghaghen *et al.*, 2009, p. 82, tradução nossa).

A ideia de gradualismo se aplica ao tratamento de textos, termos e palavras-chave, assim como ao ranqueamento de documentos recuperados em SRI. A ideia central é que "a relevância de um texto com respeito a um grupo de palavras-chave é também uma noção difusa" (Bouidghaghen *et al.*, 2009, p. 82, tradução nossa). A lógica difusa tem tido tantas aplicações na computação e nos sistemas de informação que o assunto conta com seus próprios periódicos. No levantamento realizado para este estudo, foram identificados dois: o *Fuzzy Sets and Systems*, ativo desde 1978, e o *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, ativo desde 1993.

O formalismo simbólico da lógica difusa é baseado em aspectos da lógica clássica, mas tem uma representação mais elaborada e complexa, que foge aos limites deste estudo. Zadrozny e Nowacka (2009), argumentam a favor da modelagem em RI a partir da lógica difusa, explicando sua utilidade nos seguintes termos:

A lógica difusa foi rapidamente reconhecida como uma conveniente ferramenta de modelagem formal na área da RI. Os conceitos de relevância de um documento com respeito a uma consulta ou a importância de uma palavra-chave para a representação de um documento/consulta prestam-se a uma interpretação baseada na lógica difusa" (Zadrozny; Nowacka, 2009, p. 2173, tradução nossa).

A representação de incerteza e de gradualismo de relevância é muito pobre nos termos da lógica clássica. Por isso, autores reconhecem a necessidade de que os SRI evoluam, conforme evoluem as demandas informacionais dos usuários e o próprio montante de informação a ser recuperada. "Modelos de recuperação da informação baseados na lógica difusa pode ser vistos como emergindo do modelo lógico básico, usualmente referido como o modelo booleano" (Zadrozny; Kacprzyk, 2009, p. 42, tradução nossa). Esta evolução em direção a sistemas difusos se desenvolve na base da lógica clássica, não como uma oposição ou rejeição a seus princípios, mas como um complemento para melhor aplicação na construção e análise dos SRI. Oussalah *et al.* (2008, p. 423, tradução nossa) comentam sobre isso que: "em anos recentes, tentativas de construir modelos de recuperação da informação como extensões de modelos booleanos tem sido propostos". Esta evolução, aqui analisada, já havia se insinuado nos resultados apresentados no Quadro 1, acima.

## 6 CRÍTICA CONTRA O FORMALISMO LÓGICO NA RI

O paradigma social centrado no usuário leva alguns estudiosos da CI a manifestar críticas em relação ao formalismo lógico na área da RI. A formulação dessas críticas, às vezes, se expressa nos termos de uma crítica ao princípio da não-contradição, ou ao princípio binário do terceiro excluído. Com base nos resultados obtidos no presente estudo, é possível responder a essa frente de críticas a partir de dois argumentos:

- i) Tanto o princípio binário quanto o princípio da não-contradição no sentido de um formalismo estrito e limitante, são aspectos que já foram superados na RI quando ela incorporou lógicas não-clássicas como linguagens formais de seus modelos.
- ii) Essa crítica é formalmente mal colocada, porque se dirige aos princípios lógicos e não aos sistemas particulares (em que se observa possibilidade de representação de multiplicidade e diversidade pelo sistema de RI).

A resposta a este criticismo contra o formalismo lógico deve levar o leitor a considerar que os sistemas da lógica podem vir a representar multiplicidade. Contudo, ao examinar o papel representado pela lógica na RI, é preciso fazer perguntas que ultrapassam o nível técnico e instrumental com que as teorias são consideradas usualmente. Isto conduz ao terreno da reflexão e da crítica conceituais, fora do contexto da construção de sistemas. Ali, entram em jogo as motivações provenientes dos estudos das Humanidades. A relação entre uma cultura tecnológica e uma cultura humanista, muitas vezes, envolve choque e conflito de pontos de vista. Conflitos deste tipo podem ter alguma relevância para a construção e a avaliação de SRI, dependendo da situação. Uma situação deste tipo é a da crítica que se baseia na noção de pluralismo lógico que, numa versão mais extrema, se formula nos termos de uma crítica contra os próprios princípios da lógica clássica.

Uma das versões desta crítica, que em certa medida está endereçada à RI, se acha em Soares, Martin e Francelin (2013). Os autores argumentam em favor de procedimentos mais abrangentes e inclusivos para classificar, organizar e recuperar a informação, especialmente no ambiente da Web. Segundo afirmam, sua proposta é contornar a lógica clássica a fim de “resgatar e reabilitar os conhecimentos ‘subalternos’” (Soares; Martin; Francelin, 2013, p. 63). Esta perspectiva de reabilitação e resgate de conhecimentos considerados subalternos é defendida, com base em uma crítica bastante incisiva contra a lógica formal, especialmente os princípios da lógica clássica.

Sua hipótese no artigo é a de que,

[...] a imposição da lógica clássica como jargão solitário na ciência documental caracteriza um monopólio epistemológico fragmentado, onde a classe dominante, através de suas imposições, dita o que é válido para o mundo, o que é científico ou verdadeiro – obrigações essas que afetam diretamente o indivíduo atual, às vezes sem que ele sequer perceba (Soares; Martin; Francelin, 2013, p. 58).

Fora o fato de que não há qualquer relação evidente entre a lógica clássica e a classe dominante, há outras questões a serem consideradas acerca desta posição dos autores. Logo de início, se poderia objetar que a lógica clássica se baseia em um conjunto de princípios e, com sua estrutura, é possível desenvolver teorias e extrair consequências. A suposição iconoclasta de que a lógica clássica é uma ferramenta ideológica não tem fundamento.

Uma via de resposta a essa crítica consiste em afirmar que não seria possível aceitar essa postulação do significado social da lógica. A estrutura da lógica clássica possui usos. Para alguns usos ela é extremamente bem-sucedida como, por exemplo, para a programação de computadores digitais, para a realização de testes de teorias científicas e a construção de diversos, mas não todos, sistemas de organização e recuperação do conhecimento. Em outros casos, conforme o uso, é preciso recorrer a outros princípios e, portanto, a outros sistemas de lógica. A linguagem combativa do artigo não deveria enganar o leitor atento. Se a lógica como ciência formal tem alguma aplicação para outras ciências, como é o caso da RI, é preciso desde o começo recusar e superar um olhar dogmático acerca da lógica.

Outro ponto a ser considerado é que fica mais demorado estudar lógicas não-clássicas, devido a sua maior complexidade, uma vez que normalmente são construídas para superar restrições e limitações dos sistemas clássicos. Lowe *et al.* (2018) discutem essa questão. Sua pesquisa compara o uso da recuperação booleana com a recuperação em linguagem natural, sob o horizonte do ensino da álgebra booleana no primeiro ano de cursos universitários nos Estados Unidos da América. Do ponto de vista dos usuários, os resultados indicam que é mais intuitivo usar a linguagem natural para realizar buscas. Além disso, seria mais interessante que a educação de usuários “focasse em questões mais complexas relacionadas com as buscas” (Lowe *et al.*, 2018, p. 531, tradução nossa). A operação lógica que os motores de busca fazem, por trás da interface acessível ao usuário, não precisa reproduzir a lógica que os usuários seguem ao formular suas consultas.

Se os SRI forem apoiados em sistemas não clássicos, a fim de tornar a recuperação mais inclusiva, será possível que a futura evolução dos SRI leve a uma estrutura cujo funcionamento, a maioria dos usuários não compreenderia. A perda de contato dos usuários com os fundamentos sobre os quais se constroem os SRI aparece como outra questão crítica reforçada, talvez involuntariamente, pelas críticas à lógica clássica de fundamentação binária. Escolhas quanto ao tipo de sistema lógico para a construção de um SRI são tomadas, sempre, por motivos inerentes à programação e aplicabilidade de tais programas.

## 7 CONSIDERAÇÕES FINAIS

A grande maioria dos SRI utilizados atualmente é construída com os blocos da lógica formal, tendo a álgebra booleana como seu ponto de partida básico. A base da RI, cujo crescimento e alcance social são considerados pelos autores aqui mencionados é, portanto, a lógica clássica. Desenvolvimentos e alternativas não-clássicas são sistemas construídos como modificações e complementos deste modelo básico, que é básico por uma boa razão. Sistemas alternativos para a RI estão atualmente em evolução. O limite para o desenvolvimento de sistemas com aspectos lógicos mais complexos é um limite empírico, prático, não uma resistência ideológica macabra, que planejasse preservar os interesses de alguma classe dominante, que se fortalece pelo emprego do princípio da não-contradição.

Os motores de busca vêm sendo desenvolvidos no formato de sistemas, que recuperam grande volume de informação, a partir de interfaces intuitivas para o usuário. Aqui reside o ponto central para a sobrevivência de modelos lógicos na RI. Tarefas criativas e reflexivas empreendidas por pessoas podem se basear em estruturas variadas de interpretação e formas complexas de classificação que a inteligência humana pode empregar. Mas, boa parte do processo de projeto, construção, implantação e avaliação de máquinas de busca, envolve utilização de computadores, que não usam a linguagem natural como os humanos. Daí a necessidade de traduzir a diversidade de sentidos da comunicação humana para uma forma precisa e exata, a preço de que seja mais restrita e limitada.

## REFERÊNCIAS

ABDULAHAD, Karam *et al.* Modeling Information Retrieval by Formal Logic: a survey. **ACM Computing Surveys**, [s.l.], v. 52, n. 1, 2019. DOI <https://dl.acm.org/doi/10.1145/3291043>.

ANTONIOU, Grigoris. A tutorial on Default Logics. **ACM Computing Surveys**, [s.l.], v. 31, n. 3, p. 337-359, 1999. DOI <https://dl.acm.org/doi/abs/10.1145/344588.344602>.

AUCHER, Guillaume. **Logic and commonsense reasoning**: lecture notes. Rennes: [s.n.], 2017. Disponível em: <https://hal.science/cel-01586568/file/LogicCommonsense-CourseNotes-2017.pdf>. Acesso em: 13 out 2023.

BAADER, Franz *et al.* **The description Logic handbook theory, implementation and applications.** Cambridge: Cambridge University Press, 2003.

BOUIDGHAGHEN, Ourdia *et al.* A fuzzy logic approach to topic extraction in texts. **International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems**, [s.l.], v. 17, [supl. 1], p. 81–112, 2009. Disponível em: [https://www.researchgate.net/publication/242636918\\_A\\_fuzzy\\_logic\\_approach\\_to\\_topic\\_extraction\\_in\\_texts](https://www.researchgate.net/publication/242636918_A_fuzzy_logic_approach_to_topic_extraction_in_texts). Acesso em: 13 out 2023.

CHIARAMELLA, Y.; CHEVALLET, J. P. About Retrieval Models and Logic. **The computer journal**, [s.l.], v. 35, n. 3, 1992. Disponível em: <https://academic.oup.com/comjnl/article/35/3/233/525612?login=false>. Acesso em: 13 out 2023.

D'OTTAVIANO, Ítala Maria Loffredo; FEITOSA, Hércules de Araújo. **Sobre a história da lógica, a lógica clássica e o surgimento das lógicas não-clássicas.** [s.l.]: [s.n.], 2003.

HUNTER, A. A default logic based framework for context-dependent reasoning with lexical knowledge. **Journal of Intelligent Information Systems**, [s.l.], v. 16, [s.n.], p. 65–87, 2001. Disponível em: <https://link.springer.com/article/10.1023/A:1008741010967>. Acesso em: 13 out 2023.

KAIBO Xu *et al.* A description logic based approach to formalizing the information bearing capability of paths in ER schemata. **International Journal of Intelligent Computing and Cybernetics**, [s.l.], v. 8, n. 3, p. 279-291, 2015. Disponível em: [https://www.researchgate.net/publication/277786962\\_A\\_Description\\_Logic\\_Based\\_Approach\\_to\\_Formalizing\\_the\\_Information\\_Bearing\\_Capability\\_of\\_Paths\\_in\\_ER\\_Schemata](https://www.researchgate.net/publication/277786962_A_Description_Logic_Based_Approach_to_Formalizing_the_Information_Bearing_Capability_of_Paths_in_ER_Schemata). Acesso em: 13 out 2023.

KRAUSE, Décio. Lógica paraconsistente. **Crítica**, [s.l.], [s.n.], [s.n.], nov. 2003. Disponível em: [https://criticanarede.com/log\\_paraconsistente.html](https://criticanarede.com/log_paraconsistente.html). Acesso em: 13 out 2023.

LALMAS, Mounia. Logical models in information retrieval: Introduction and overview. **Information Processing & Management**, [s.l.], v. 34, n. 1, p. 19-33, 1998. Disponível em: <https://www.sciencedirect.com/journal/information-processing-and-management/vol/34/issue/1>. Acesso em: 13 out 2023.

LOWE, M. Sara *et al.* The Boolean is dead, long live the Boolean! Natural language versus Boolean searching in introductory undergraduate instruction. **College & Research Libraries**, [s.l.], v. 79, n. 4, p. 517-534, 2018. Disponível em: <https://crl.acrl.org/index.php/crl/article/view/16729>. Acesso em: 13 out 2023.

MAHR, Bernt. Information science and the logic of models. **Software System Models**, v. 8, [s.n.], p. 365–383, 2009. Disponível em: <https://link.springer.com/article/10.1007/s10270-009-0119-2>. Acesso em: 13 out 2023.

MANNING, Christopher D.; RAGHAVAN, Prabhakar; SCHÜTZE, Hinrich. **An Introduction to Information Retrieval**. Cambridge: Cambridge University Press. 2009.

OUSSALAH *et al.* Personalized information retrieval system in the framework of fuzzy logic. **Expert Systems with Applications**, [s.l.], [s.n.], n. 35, p. 423–433, 2008. Disponível em:

[https://www.researchgate.net/publication/220215007\\_Personalized\\_information\\_retrieval\\_system\\_in\\_the\\_framework\\_of\\_fuzzy\\_logic/link/53f0e9840cf2711e0c43153b/download](https://www.researchgate.net/publication/220215007_Personalized_information_retrieval_system_in_the_framework_of_fuzzy_logic/link/53f0e9840cf2711e0c43153b/download). Acesso em: 13 out 2023.

REITER, R. A logic for default reasoning. **Artificial Intelligence**, [s.l.], v. 13, n. 1-2, p. 81-132, Apr. 1980. DOI <https://dl.acm.org/doi/10.1016/0004-3702%2880%2990014-4>.

SEBASTIANI, Fabrizio. On the role of logic in information retrieval. **Information Pressing & Management**, [s.l.], v. 34, n. 1, p. 1-18, 1998. Disponível em: <https://www.sciencedirect.com/science/article/abs/pii/S0306457397000551>. Acesso em: 13 out 2023.

SEBASTIANI, Fabrizio. Towards a logical reconstruction of information retrieval theory. **Cybernetics & Systems**, [s.l.], v. 30, n. 5, p. 411-428, 1999. Disponível em: [https://www.researchgate.net/publication/2614160\\_Towards\\_a\\_Logical\\_Reconstruction\\_of\\_Information\\_Retrieval\\_Theory/link/5510e1a90cf20352196cc96f/download](https://www.researchgate.net/publication/2614160_Towards_a_Logical_Reconstruction_of_Information_Retrieval_Theory/link/5510e1a90cf20352196cc96f/download). Acesso em: 13 out 2023.

SOARES; MARTIN; FRANCELIN. Pluralismo lógico e epistemografia interativa como ferramentas desclassificadoras do conhecimento. **Rev. digit. bibliotecon. cienc. inf.** Campinas, v. 11, n. 1, p. 55-71. 2013. Disponível em: <https://periodicos.sbu.unicamp.br/ojs/index.php/rdbci/article/view/1651>. Acesso em: 13 out 2023.

VAN RIJSBERGEN, Cornelis Joost. A non-classical logic for information retrieval. **The Computer Journal**, [s.l.], [s.n.], n. 29, p. 481-485, 1986.

ZADROŻNY, Sławomir; KACPRZYK, Janusz. On an interpretation of keywords weights in information retrieval: some fuzzy logic based approaches. **International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems**, [s.l.], v. 17, [supl. 1], p. 41-58, 2009. DOI <https://doi.org/10.1142/S0218488509006029>.

ZADROŻNY Sławomir; NOWACKA, Katarzyna. Fuzzy information retrieval model revisited. **Fuzzy Sets and Systems**, [s.l.], v. 160, n. 15, p. 2173–2191, Aug. 2009. Disponível em: <https://www.sciencedirect.com/science/article/abs/pii/S0165011409001080>. Acesso em 13 out 2023.