



UNIVERSIDADE ESTADUAL PAULISTA
“JÚLIO DE MESQUITA FILHO”
Campus de Marília

RICARDO CÉSAR DE CARVALHO

CHATBOT APLICADO À RECUPERAÇÃO DE INFORMAÇÃO:
Um modelo orientado a metadados

MARÍLIA
2022

RICARDO CÉSAR DE CARVALHO

CHATBOT APLICADO À RECUPERAÇÃO DE INFORMAÇÃO:
Um modelo orientado a metadados

Tese apresentada ao Programa de Pós-Graduação em Ciência da Informação como parte das exigências para a obtenção do título de Doutor em Ciência da Informação pela Faculdade de Filosofia e Ciências, Universidade Estadual Paulista (UNESP), Campus de Marília.

Área de concentração: Informação, Tecnologia e Conhecimento.

Linha de pesquisa: Informação e Tecnologia.

Orientador: Prof. Dr. Edberto Ferneda

MARÍLIA
2022

C331c Carvalho, Ricardo César de
Chatbot aplicado à recuperação de informação : um modelo
orientado a metadados / Ricardo César de Carvalho. -- Marília,
2022
214 f.

Tese (doutorado) - Universidade Estadual Paulista (Unesp),
Faculdade de Filosofia e Ciências, Marília
Orientador: Edberto Ferneda

1. Ciência da Informação. 2. Recuperação da Informação. 3.
Metadados. 4. Interfaces de Usuário. 5. Chatbots. I. Título.

Sistema de geração automática de fichas catalográficas da Unesp. Biblioteca da
Faculdade de Filosofia e Ciências, Marília. Dados fornecidos pelo autor(a).

Essa ficha não pode ser modificada.

RICARDO CÉSAR DE CARVALHO

CHATBOT APLICADO À RECUPERAÇÃO DE INFORMAÇÃO:
Um modelo orientado a metadados

Tese apresentada ao Programa de Pós-Graduação em Ciência da Informação da Universidade Estadual Paulista “Júlio de Mesquita Filho” (Unesp), como requisito parcial para a obtenção do título de Doutor em Ciência da Informação.

Área de Concentração: Informação, Tecnologia e Conhecimento.
Linha de Pesquisa: Informação e Tecnologia.

Banca Examinadora

Prof. Dr. EDBERTO FERNEDA (Orientador)

Departamento de Ciência da Informação

Universidade Estadual Paulista Júlio de Mesquita Filho (UNESP)

Prof. Dr. CECILIO MERLOTTI RODAS

Departamento de Ciência da Informação

Universidade Estadual Paulista Júlio de Mesquita Filho (UNESP)

Prof. Dr. CARLOS CANDIDO DE ALMEIDA

Departamento de Ciência da Informação

Universidade Estadual Paulista Júlio de Mesquita Filho (UNESP)

Profa. Dra. LUCIANA MARIA VIEIRA PÖTTKER

Departamento de Computação

Instituto Federal de Educação, Ciência e Tecnologia do Paraná (IFPR)

Prof. Dr. GUILHERME ATAÍDE DIAS

Departamento de Ciência da Informação

Universidade Federal da Paraíba (UFPB)

Marília, 14 de setembro de 2022.

Dedico este trabalho

À Priscila, minha esposa, o meu presente,

À Laís e Lisa, minhas filhas, o meu futuro,

À Neuza, minha mãe e Adélio, meu pai (*In Memoriam*), o meu passado.

AGRADECIMENTOS

Agradeço primeiramente à **Deus**, por me dar sabedoria e perseverança para superar todas as adversidades que aconteceram no percurso deste trabalho.

À minha esposa **Priscila**, pela paciência nos inúmeros momentos que não pude ajudar ou estar presente. Pelo apoio e incentivo diário que me deram mais energia para realizar este trabalho. Sem você não seria possível, nem tampouco, existiria razão para fazê-lo. Te amo.

À minha filha **Laís**, por todos os momentos que passou ao meu lado, atrás de mim, debaixo da minha cadeira ou no meu colo, tão pequena, não conhece as razões para algo assim, mas sabia que era muito importante e sempre me perguntava se após o meu “trabalho” eu poderia lhe dar alguma atenção, com certeza eu estava ali e agora ainda mais.

À minha filha **Lisa**, que veio ao mundo no meio da pandemia, muito diferente da sua irmã, super curiosa, gosta de me assistir a trabalhar e ficar no meu colo para me “ajudar” a teclar, adormecendo abraçadinha em mim ao se cansar.

Aos meus pais, **Adélio e Neuza**, que sempre estiveram ao meu lado, me incentivando a vida toda a estudar e me dedicar, mesmo não entendendo o meu estudo, não pouparam esforços para que nossa família sempre tivesse o que era necessário. Pai, você nos deixou na reta final deste trabalho, nem tivemos a oportunidade de nos despedir, mas onde estiver, espero que esteja orgulhoso.

Ao meu orientador, **Prof. Dr. Edberto Ferneda**, que desde antes de ser discente da UNESP, me aceitou na disciplina e apresentou a área da CI e da RI, tornando-se um bom amigo. Nos momentos de conversa durante os almoços no shopping, ou mesmo no campus, sempre faz o que pode para auxiliar, orientar, incentivar ou desconstrair seus alunos. Agradeço imensamente por ter me aceito como aluno orientado, me perdoe os sumiços neste período, obrigado por me

ajudar nas correções e na orientação desse trabalho, e ter mudado a minha vida, pois isso nunca será esquecido.

Aos professores **Dr. Cecílio Merlotti Rodas** e **Dra. Luciana Maria Vieira Pöttker**, por aceitarem ao convite de participar da minha banca de qualificação e por todas as contribuições que foram direcionadas para a melhoria do meu trabalho.

Aos professores **Dr. Carlos Candido De Almeida** e **Dr. Guilherme Ataíde Dias**, por terem aceitado integrar a minha banca de defesa da tese, pelo seu tempo na leitura e contribuições neste trabalho.

Aos **professores, colegas e amigos do Programa de Pós-Graduação em Ciência da Informação da UNESP – Campus Marília**, com os quais pude aprender muito, pois a diversidade de conhecimento presente em uma aula, grupo de estudo ou conversa de corredor, estimula a capacidade de entender, aprender e questionar e no futuro, melhor ensinar.

Ao **Instituto Federal de São Paulo (IFSP)**, pelo programa de afastamento para qualificação, que permitiu a minha dedicação em tempo integral, na segunda metade do período do doutorado, na construção do conhecimento presente neste trabalho e, assim me tornar um professor melhor para a educação dos nossos alunos.

Aos **professores e servidores do Instituto Federal de São Paulo (IFSP) – Campus Votuporanga**, pelo apoio em todos os momentos que precisei de auxílio enquanto estive afastado para a realização deste trabalho.

E por fim, e não menos importante, agradeço a todos que de alguma maneira direta ou indiretamente contribuíram na construção deste trabalho, desde a ideia até a conclusão.

Muito Obrigado!

“E você. É Jerry Shaw. Compras, preferências e dados quantificáveis que definimos como sua personalidade. Nós monitoramos todas as redes sociais, blogs, torpedos e textos eletrônicos, seus conhecidos, amigos, colegas, e-mails recebidos e enviados, uso de celular. Onde também utilizamos câmeras de segurança e de tráfego para analisar seus movimentos. Nós usamos esses dados para formar os perfis de personalidade. Nós sabemos quem vocês são. Nós estamos em toda parte.”

Fala do supercomputador autônomo Analista de Integração de Inteligência de Reconhecimento Autônomo (ARIIA), personagem do filme Controle Absoluto (2008).

RESUMO

Com a disseminação do uso da Internet e o aumento da quantidade de informação disponível para utilização pelos mais diversos tipos de pessoas, ferramentas que amparem este processo devem estar disponíveis e acompanhar o avanço da tecnologia. Neste contexto, as pesquisas por métodos de recuperação de informação se destacam pela experiência de décadas e podem auxiliar os profissionais de diversas áreas a organizar os repositórios e fornecer meios tanto para a criação de novas informações, quanto para a manutenção, armazenamento e recuperação destas. A utilização de novas interfaces, como *chatbots*, permite ao usuário, por meio de aplicativos de mensagens que já utiliza, acesso a informações na Web, sem configurações especiais e utilizando linguagem natural. Com o propósito de verificar a capacidade de um *chatbot* na função de um sistema de recuperação de informação, na busca de qualidade dos resultados e um menor índice de problemas de comunicação, propõe-se um modelo conceitual de recuperação de informação por meio de um *chatbot* disponível em aplicativos de mensagens, ou redes sociais. Devido à recuperação de informação ser um processo comunicacional, ruídos terminológicos podem se tornar um problema. A proposta para minimizar isso é o enriquecimento dos termos utilizados pelo usuário nos questionamentos, a partir de termos provindos dos metadados da base de dados de um repositório. A natureza aplicada desse trabalho segue os preceitos de uma pesquisa exploratória, podendo ser dividida em duas fases: a pesquisa bibliográfica e a elaboração de um modelo. Por meio de uma pesquisa bibliográfica em fontes primárias de informação para determinar os aspectos teóricos e embasamento a respeito da intersecção das áreas de *Chatbots*, Recuperação de Informação e Metadados. A outra fase trata-se da proposta de um modelo para atender esta pesquisa que se encontra dentro da Ciência da Informação e mais especificamente na Recuperação de Informação, com a proposta de apresentar um sistema de recuperação de informação mais eficiente por meio de uma nova interface e com acesso a metadados. Os resultados apresentados ao utilizar metadados em métodos de recuperação de informação, utilizando-os nos processos comunicacionais entre um usuário e o sistema do *chatbot*, permite constatar que a presença destes metadados pode ser um diferencial para o aumento da qualidade dos resultados e uma possibilidade para resolver a subjetividade e imprecisão dos diálogos. Conclui-se que um sistema de recuperação de informação é mais eficiente ao utilizar metadados em interfaces de agentes conversacionais, permitindo aos usuários maior inclusão de pessoas, melhores resultados, com menor número de interações e menor utilização de tecnologias complexas de análise de dados.

Palavras-Chave: Ciência da Informação; Recuperação da Informação; Metadados; Interfaces de Usuário; Chatbots.

ABSTRACT

With the widespread use of the Internet and the information amount increasement available for use by different people, tools that support this process must be available and keep up with the advancement of technology. In this context, research on information retrieval methods stands out for its decades of experience and can help professionals from different areas to organize repositories and provide means for both the creation of new information and for their maintenance, storage and retrieval. The use of new interfaces, such as chatbots, it allows the user, through messaging applications that he already knows, to access information on the Web, without special configurations and using natural language. With the purpose of verifying the capacity of a chatbot in the function of an information retrieval system, in the search for quality of results and a lower rate of communication problems, is proposed a conceptual model of information retrieval through of a chatbot available in messaging apps, or social networks. To information retrieval being a communication process, terminological noise can become a problem. The proposal to minimize this is to enrich the terms used by the user in the inquiries, from terms coming from the metadata of the repository's database. The applied nature of this work follows the precepts of an exploratory research, which can be divided into two phases: the bibliographic research and the elaboration of a model. Through bibliographic research in primary sources of information to determine the theoretical aspects and basis regarding the intersection of the areas of Chatbots, Information Retrieval and Metadata. The other phase is the proposal of a model to meet this research that is within Information Science and more specifically in Information Retrieval, with the proposal to present a more efficient information retrieval system through a new interface and access to metadata. The results presented when using metadata in information retrieval methods, using them in the communication processes between a user and the chatbot system, show that the presence of this metadata can be a differential for increasing the quality of the results and a possibility to solve the subjectivity and imprecision of the dialogues. It is concluded that an information retrieval system is more efficient when using metadata in conversational agent interfaces, allowing users to include more people, better results, with fewer interactions and less use of complex data analysis technologies.

Keywords: Information Science; Information Retrieval; Metadata; User Interfaces; Chatbots.

LISTA DE FIGURAS

Figura 1 – Pesquisas dos termos <i>chatbot</i> e <i>chatterbot</i> no Mundo	31
Figura 2 – Pesquisas dos termos <i>chatbot</i> e <i>chatterbot</i> no Brasil	31
Figura 3 - Representação do processo de recuperação de informação	40
Figura 4 - Uso de dispositivos móveis e <i>desktop</i> no mundo de Jan 2009 a Ago 2020	59
Figura 5 - Uso de dispositivos móveis e <i>desktop</i> no Brasil de Jan 2009 a Ago 2020	60
Figura 6 - Módulos de uma Interface Conversacional	78
Figura 7 - Cinco níveis de assistentes de IA.....	80
Figura 8 – Google Duplex como módulo do Assistente.....	82
Figura 9 – Exemplo de catálogo MARC	90
Figura 10 – Exemplo de <i>rich snippets</i>	92
Figura 11 – Página de busca do catálogo Athena	96
Figura 12 – Página de Busca do BENANCIB.....	98
Figura 13 – Página de Busca do BDTD	100
Figura 14 – Exemplo de diálogo entre um humano e um <i>chatbot</i>	104
Figura 15 – Página de pesquisa do ENANCIB WEB.....	115
Figura 16 – Diagrama do Modelo.....	117
Figura 17 – Utilização do buscador do sítio ENANCIB WEB	118
Figura 18 – Interfaces conversacionais do ENANCIB WEB.....	119
Figura 19 – Módulos de uma interface conversacional.....	120
Figura 20 – Proposta de API.....	122
Figura 21 – Compartilhamento de um mesmo repositório por múltiplas interfaces	123
Figura 22 – Agentes no Dialogflow.....	125

Figura 23 – Default Welcome Intent (Intenção Padrão de Boas-Vindas).....	126
Figura 24 – A função da <i>Intent</i>	127
Figura 25 – <i>Fulfillment</i> em funcionamento	128
Figura 26 – <i>Fulfillment</i> – Acesso a servidores externos.....	129
Figura 27 – <i>Fulfillment</i> – Desenvolvimento interno	130
Figura 28 – Integrações disponíveis	131
Figura 29 – Utilização da Intenção e sua resposta.....	133
Figura 30 – Utilização de Entidades e sua resposta.....	135
Figura 31 – Utilização de Contexto e sua resposta.....	136

LISTA DE QUADROS

Quadro 1 – Resultado das pesquisas na BDTD.....	27
Quadro 2 – Resultado das pesquisas na NDLTD.....	28
Quadro 3 – As plataformas, módulos e tecnologias que compõem um <i>chatbot</i>	79
Quadro 4 – Metadados exportados pela página da ATHENA	96
Quadro 5 – Metadados extraídos da página do BENANCIB	98
Quadro 6 – Metadados extraídos da página do BDTD.....	101
Quadro 7 – Diálogo entre usuário e <i>chatbot</i> fictício da BDTD.....	107
Quadro 8 – Informações das Edições do ENANCIB	112
Quadro 9 – Módulos e Ações da Interface Conversacional	120
Quadro 10 – Diálogo entre usuário e <i>chatbot</i> do Projeto	137
Quadro 11 – Intenções	147
Quadro 12 – Entidades	147
Quadro 13 – Contextos	149
Quadro 14 – Proposta de Avaliação por Nielsen	154

LISTA DE ABREVIATURAS E SIGLAS

AACR	<i>Anglo-American Cataloguing Rules</i>
ACM	<i>Association of Computing Machinery</i>
AIML	<i>Artificial Intelligence Markup Language</i>
ANCIB	Associação de Pesquisa e Pós-graduação em Ciência da Informação
API	<i>Application Programming Interface</i>
BDTD	Biblioteca Digital Brasileira de Teses e Dissertações
CI	Ciência da Informação
CLN	Compreensão de Linguagem Natural
DC	<i>Dublin Core (Dublin Metadata Core Element Set)</i>
DTD	<i>Document Type Definition</i>
ENANCIB	Encontros Nacionais de Pesquisa e Pós-Graduação em Ciência da Informação
ETD-MS	<i>Electronic Thesis and Dissertations Metadata Standard</i>
GLN	Geração de Linguagem Natural
HTML	<i>HyperText Markup Language</i>
HTTP	<i>Hypertext Transfer Protocol</i>
IBM	<i>International Business Machines</i>
IHC	Interação Humano-Computador
IoT	<i>Internet of Things</i>
LC	<i>Library of Congress</i>
MARC	<i>Machine Readable Cataloging</i>

METS	<i>Metadata Encoding and Transmission Standard</i>
MRI	Modelo de Recuperação de Informação
MTD-BR	Metadados para Descrição de Teses e Dissertações
NDLTD	<i>Networked Digital Library of Theses and Dissertations</i>
NISO	Organização Nacional de Padrões de Informação
NoSQL	<i>Not Only SQL</i>
OAIS	<i>Open Archival Information System</i>
OCLC	<i>Online Computer Library Center</i>
PLN	Processamento de Linguagem Natural
RDF	<i>Resource Description Framework</i>
SAC	Serviço de Atendimento ao Consumidor
SMS	<i>Short Message Service</i>
SQL	<i>Structured Query Language</i>
SRI	Sistema de Recuperação de Informação
UFF	Universidade Federal Fluminense
UNESP	Universidade Estadual Paulista
URI	<i>Uniform Resource Identifier</i>
URL	<i>Uniform Resource Locator</i>
XML	<i>Extensible Markup Language</i>

SUMÁRIO

1. INTRODUÇÃO	16
1.1 Caracterização do problema de pesquisa.....	21
1.2 Hipóteses	22
1.3 Tese	23
1.4 Proposta da pesquisa	23
1.5 Objetivos	23
1.5.1 Objetivo geral	24
1.5.2 Objetivos específicos.....	24
1.6 Justificativa.....	25
1.7 Procedimentos Metodológicos	29
1.8 Terminologia Adotada.....	30
1.9 Organização do trabalho.....	32
2. RECUPERAÇÃO DE INFORMAÇÃO	34
2.1 Recuperação de Informação como Área de Pesquisa.....	36
2.2 O Processo de Recuperação de Informação	40
2.2.1 Documentos (corpus).....	40
2.2.2 Representação dos documentos	41
2.2.3 Usuário	42
2.2.4 Expressão de Busca	42
2.2.5 Representação da Expressão de busca.....	42
2.2.6 Função de busca	43
2.2.7 Resultado da busca	43
2.2.8 Modelo de Recuperação de Informação	43
2.3 Processamento da Linguagem Natural.....	46
2.3.1 Normalização de variações linguísticas.....	47
2.3.2 Identificação de termos compostos.....	48
2.3.3 Resolução de ambiguidade	48
2.4 Recuperação de Informação como um Processo de Comunicação	49

3. CHATBOTS	55
3.1 Definição de <i>Chatbot</i>	55
3.2 Características dos <i>chatbots</i>	60
3.3 Tipos de <i>Chatbots</i>	62
3.4 Diferença entre Assistente Digital e <i>Chatbot</i>	65
3.5 Plataformas Conversacionais	66
3.5.1 Plataformas de conversação.....	66
3.5.2 Assistentes virtuais	69
3.5.3 Ferramentas para construção de <i>chatbots</i>	71
3.5.4 Persistência de Dados	74
3.5.5 Métricas de Avaliação	75
3.6 Módulos de um <i>Chatbot</i>	77
4. METADADOS	83
4.1 Definição de Metadados	84
4.2 Padrões de Metadados	88
4.3 Uso de Metadados	91
5. MODELO PROPOSTO	103
5.1 Modelo de Interface Conversacional.....	116
5.1.1 Dialogflow	123
5.1.2 ENANCIB AGENTE – <i>Chatbot</i> conceitual	132
5.2 Avaliação do Modelo e Interface	152
6. CONSIDERAÇÕES FINAIS.....	156
ANEXO I	169
ANEXO II.....	184
ANEXO III	200

1. INTRODUÇÃO

Nos últimos anos nunca se produziu tantas informações, o que permitiu que diversas áreas do conhecimento se beneficiassem dessa explosão informacional e buscassem meios para usufruir de tal acontecimento. Um dos fatores que corroborou com esta situação foi o surgimento da Web, descrita por seu criador, Tim Berners-Lee (2017), como uma “plataforma aberta que permitiria a todo mundo e em todos os lugares compartilhar informações, ter oportunidades de acesso e colaborar através de barreiras geográficas e culturais”¹.

Inicialmente, apenas entidades governamentais puderam ter acesso a esta grande rede para o compartilhamento do conhecimento, mas nas últimas décadas, essa tecnologia chegou às residências das pessoas, por meio dos mais diferentes dispositivos: como computadores, *laptops*, *tablets*, celulares, videogames, TVs, carros, entre outros; o que criou muitos desafios para a criação, manutenção, disponibilização e acesso a informações (CETIC.BR, 2020b).

Para produzir conteúdo para a Web, como na criação de um repositório de informações digitais, o seu proprietário tem a opção de escolher como essas informações serão descritas e armazenadas, conseqüentemente, determinar como elas serão acessadas e recuperadas, disponibilizando interfaces apropriadas. As informações podem ser enriquecidas atribuindo-lhes um conjunto de metadados, e permitindo ao usuário utilizá-los com o propósito de auxiliá-lo durante o processo de recuperação de informação.

Recuperar informação consiste em identificar em um conjunto de documentos aqueles que satisfaçam uma determinada necessidade de informação. Um sistema de recuperação de informação é um elemento mediador nesse processo, que envolve, por um lado, um acervo documental onde cada documento é representado por expressões linguísticas que resumem o

¹ [...] open platform that would allow everyone, everywhere to share information, access opportunities, and collaborate across geographic and cultural boundaries.

seu conteúdo informacional. Por outro lado, os utilizadores desse sistema tentam descrever linguisticamente as suas necessidades de informação a fim de obterem documentos que venham a satisfazer tais necessidades. A recuperação de informação se realiza por meio da comparação entre a representação de cada documento e a representação da necessidade de informação do usuário. Portanto, um sistema de recuperação de informação é um ambiente linguístico mediador da comunicação entre um estoque de informação e seus requisitantes (FERNEDA, 2013).

A recuperação de informação pode ser vista como um processo de comunicação. Segundo Meadow *et al* (2007, p.3), a recuperação de informação é um meio pelo qual autores e criadores de registros se comunicam com os leitores. Para Vieira (1994, p. 6), a recuperação de informação é um processo em que emissor e receptor interagem para atender a uma necessidade de informação. "Essa interação só é viável por meio do uso da linguagem".

Segundo Crestani e Pasi (2003, p. 162, tradução nossa):

A subjetividade é uma propriedade intrínseca de qualquer sistema de RI. Está relacionado ao próprio conceito de relevância. É um fato bem conhecido que o mesmo documento pode ser totalmente relevante para um usuário e totalmente irrelevante para outro usuário, embora ambos façam a mesma consulta ao mesmo sistema de RI. Somente o usuário é o juiz final da relevância de um documento para uma necessidade de informação.²

Assim, a área da Recuperação de Informação foi criada para pesquisar metodologias que descrevem técnicas de como extrair informações a partir de uma grande quantidade de documentos. Para tal, são necessárias diversas habilidades no tratamento e recuperação das informações, habilidades estas que podem ser encontradas em diferentes tipos de profissionais, em diversas áreas de conhecimento.

A partir de 1960, na investigação de como as informações poderiam se comunicar de maneira mais eficiente, por meio da discussão entre pesquisadores, especialistas ou técnicos de vários campos do conhecimento, a saber, bibliotecários, matemáticos, físicos, biólogos, médicos e químicos. Como resultado deste trabalho, emergem diversos temas a respeito das tecnologias daquela época que começavam a ser exploradas com o intuito de atender às necessidades informacionais dos setores científicos, tecnológicos e industriais, para resolver um problema que já começava a aparecer, a grande quantidade de informações geradas e o

² Subjectivity is an intrinsic property of any IR system. It is related to the concept of relevance itself. It is a well known fact that the same document may be fully relevant to a user and fully not relevant to another user, although both posing the same query to the same IR system. Only the user is the final judge of the relevance of a document to an information need.

devido acesso a elas (ALVARES; ARAÚJO JÚNIOR, 2010). Durante as discussões sobre esses temas, o conceito da Ciência da Informação foi criado, denominaram-na como uma ciência que:

(...) investiga as propriedades e o comportamento da informação, as forças que governam o fluxo da informação e os meios de processamento da informação para acessibilidade e usabilidade ótimas. Os processos incluem a geração, disseminação, coleta, organização, armazenamento, recuperação, interpretação e uso da informação. A área é derivada de ou relacionada à matemática, lógica, linguística, psicologia, tecnologia computacional, pesquisa operacional, artes gráficas, comunicações, biblioteconomia, administração e algumas outras áreas (NATIONAL SCIENCE FOUNDATION, 1962, p. 264-265).

Como essa ciência é derivada da convergência de várias outras áreas, pode ser considerada interdisciplinar, pois sendo a informação seu objeto de estudo, independente de tipo ou natureza. Como descrito por Borko (1968, p. 3), a Ciência da Informação “tem componentes de ciência pura quando questiona o assunto sem se preocupar com sua aplicação, e componentes de ciência aplicada quando desenvolve serviços e produtos.”

Ainda sobre a interdisciplinaridade da Ciência da Informação, Saracevic (1996, p.48) descreve quatro ciências que mantêm relações estreitas com ela e abrangem muitos profissionais que podem atuar conjuntamente, entre elas, a Biblioteconomia, a Ciência da Computação, a Ciência Cognitiva e a Comunicação. A respeito da interseção entre a Ciência da Informação e a Ciência da Computação, Lima (2003), discute que por meio da representação da informação, sua organização intelectual e encadeamentos como a busca e recuperação de informação, a qualidade, o valor e uso da informação, estas ciências podem contribuir muito uma com a outra.

A partir da visão da qual a Ciência da Informação pode ser beneficiada das diferentes formas de interação do usuário ao se dedicar a sistemas com interfaces inteligentes, que utilizam novas formas de se comunicar com um usuário, além de uma investigação a respeito da disciplina de Interação Humano-Computador, podemos referir a respeito das contribuições das Ciências Cognitivas. Sobre isso, Lima (2003, p.81) também complementa que a Ciência Cognitiva cresceu a partir de 3 pontos:

- Desenvolvimento da psicologia do processamento da informação, na qual a meta era especificar o processamento interno envolvido na percepção, linguagem, memória e pensamento;
- A invenção dos computadores e as tentativas de projetar programas que pudessem fazer tarefas que as pessoas fazem;
- Desenvolvimento da teoria da gramática generativa e outras derivações da linguística.

Prosseguindo, Lima cita que o processo cognitivo envolve atividades mentais como o pensamento, a imaginação, a recordação, a solução de problemas, a percepção, o julgamento, a aprendizagem da linguagem, entre outras. Ao se buscar a convergência das duas ciências, Ciência da Informação e Ciência da Computação, Marc de Mey (1992, p.4) afirmou que “o ponto de vista cognitivo da Ciência da Informação implica que cada ato de processamento da informação, seja ele perceptivo ou simbólico, é mediado por um sistema de categorias e conceitos os quais, para o mecanismo de processamento da informação, constituem um modelo de mundo.” Ainda sobre isso, Lima (2003) questiona se o problema principal da pesquisa cognitiva em Ciência da Informação poderia ser apresentado na seguinte questão: “De que maneira as estruturas conceituais que formam o universo do usuário (enquanto processador da informação) têm correspondência com a estrutura conceitual que forma o universo do sistema de recuperação?”.

Ao buscar novas formas de interação que sejam mais eficientes, processos que utilizem conceitos cognitivos podem ajudar a tornar esse “mundo” das informações mais próximo do “mundo” das necessidades dos usuários, principalmente se essa interação for transparente ou natural, como a interação entre as pessoas. A respeito disso, Jacob e Shaw (1998) observam que a ênfase em sistemas amigáveis para o usuário, com interfaces inteligentes, caracteriza as tendências de pesquisas cognitivas em ciência da informação.

Ao se referir a interface, Dias (1994, p.1) define-a como “uma superfície de contato com a informação e um envelope para o conteúdo, procurando-se adequar esta superfície aos fatores humanos envolvidos no processo de contato e às normas da organização da informação”. A partir do que foi apontado até aqui, é possível visualizar diversas possibilidades de pesquisas para se adequar a forma do acesso a sistemas de informação seguindo os preceitos da Ciência da Informação, da Ciência da Computação e da Ciência Cognitiva.

Atualmente, uma série de novos produtos e serviços passaram a ser disponibilizados por grandes empresas da computação, como Google, IBM, Microsoft, entre outras, na área de interfaces para acesso a informações, e podem ser encontradas descritas como ferramentas de computação cognitiva, que é um grande avanço da Inteligência Artificial, que utiliza de conceitos da Ciência Cognitiva, da Ciência da Computação e da Ciência da Informação para interagir com os usuários para os mais diferentes usos.

Para apresentar o comportamento de um sistema cognitivo e das diversas possibilidades, Watson (2018) descreve, que nos próximos anos, serão disponibilizadas ferramentas que

poderão interagir com os humanos e mudar o modo de como estes usam um sistema computacional, como

- Entender, praticamente, todos os tipos de dados, incluindo linguagem natural, imagens, sons e outros dados não estruturados.
- Fazer deduções fundamentadas, incluindo gerar e testar hipóteses.
- Aprender com interações contínuas, que são revisadas por especialistas e devolvidas ao sistema para ajudá-lo a interpretar melhor as informações.
- Interagir naturalmente com seres humanos, comunicando-se em linguagem natural e transmitindo conclusões que sejam úteis e compreensíveis.

Ainda complementa que um sistema cognitivo deste tipo vai além da inteligência artificial. Ele combina capacidades de inteligência artificial, aprendizado de máquina e comunicação.

Como todo e qualquer sistema comunicacional, o processo de recuperar informações detém os mesmos problemas inerentes ao processo comunicacional, como a subjetividade, a imprecisão e a ambiguidade. Segundo Alves e Santos (2013), os processos comunicacionais desempenhados pelos usuários na busca das informações nos catálogos bibliográficos, ou em bancos de dados, são os atos de encontrar, identificar, selecionar, adquirir e navegar nessas informações, não importando se o meio do recurso é físico ou digital. Nesta pesquisa, a principal contribuição é a utilização dos metadados na tentativa de solucionar alguns desses problemas em sistemas digitais, pois, espera-se através deles a representação, individualização, intercâmbio, interoperabilidade entre sistemas e o acesso e a recuperação destes recursos para qualquer informação.

Somente com a representação adequada e padronizada dos recursos informacionais, através de metadados, é garantida a maior precisão na recuperação e tratamento da informação. Existe uma grande variedade de tipos de metadados separados para cada domínio onde há uma diversidade de representações, como exemplo, os aplicados aos acervos das bibliotecas para seu controle.

De um modo geral, a principal função de um metadado seria garantir o gerenciamento da informação em um sistema, de forma que sua identificação, descrição e controle sejam o principal propósito da aplicação (ALVES; SANTOS, 2013).

Assim, existem dados disponibilizados e a necessidade dos usuários de encontrar o que precisam, a tecnologia que permite tal conexão já é especificada pela Recuperação de Informação; a Ciência da Informação contribui para a criação de dados mais ricos e a Ciência

da Computação prove os meios computacionais para tal. Como meio de acesso às informações de forma interativa, podem ser definidos na forma de sistemas cognitivos que, interagindo com os usuários, podem exigir sistemas com maior ou menor grau de inteligência e complexidade.

Hoje em dia, um número cada vez maior de usuários utiliza aparelhos celulares e outros dispositivos para acessar a Internet, e por meio destes, acessam todo tipo de informação nos mais variados repositórios e formatos. A novidade e a relevância desses novos meios de acesso a informações, baseada na possibilidade de aplicar essa tecnologia na recuperação de informações com foco na melhoria da relevância dos resultados, por meio de novas interfaces comunicacionais, constituem a motivação para o desenvolvimento de pesquisas nesta área.

1.1 Caracterização do problema de pesquisa

Considerando que uma grande parte da população já tem acesso à Internet (CETIC.BR, 2022) e o faz por diferentes tipos de dispositivos, sistemas de recuperação de informação em diferentes formatos já existem há tempos e são utilizados por variados níveis de usuários. Além disso, uma série de novas aplicações surgiram com a possibilidade da utilização destes novos dispositivos e em praticamente qualquer lugar, como as redes sociais e os sistemas de mensagens instantâneas, e neste momento novas formas de interfaces passaram a ser discutidas em muitas áreas, e por muitas empresas, com o propósito de contribuir com estas aplicações (Watson, 2018), inclusive na Ciência da Informação, na busca de aplicá-las em novos e diferentes contextos (KAUSHIK, 2019).

Diante disso, ao analisar como a Ciência da Informação aplica as técnicas de recuperação de informação por meio de uma interface conversacional com processamento de linguagem natural, verificou-se que existem muitos trabalhos nessa área, mas pouco foi escrito a respeito de como auxiliar o usuário durante o processo de escolha dos termos ao interagir para recuperar informações, vide Quadro 1 adiante. Ainda sobre as informações, diversos repositórios armazenam uma quantidade de informações estruturadas na forma de metadados e ainda permitem a busca dentro desse conteúdo, como por exemplo, o repositório da Biblioteca Digital Brasileira de Teses e Dissertações³ (BDTD), mas ainda não disponibilizam uma interface em linguagem natural para ser acessada por meio de uma rede social ou aplicativo de mensagens.

³ Disponível em: <https://bdtd.ibict.br/vufind>. Acesso em: 21 ago. 2021.

Além destes fatos, a experiência profissional do autor traz informações a respeito da presença de informações importantes e da necessidade de recuperá-las pelo usuário, ao se tratar de dados médicos de pacientes, sempre houve questionamentos sobre novas formas naturais para recuperar informações de prontuários de maneira rápida ou sob estresse, pois um médico precisa às vezes ter acesso a informações e não tem tempo para utilizar as mãos para digitar, e uma interface que pudesse acompanhá-lo durante um procedimento poderia auxiliá-lo no diagnóstico, pois as informações que ele necessita estão lá, só não podem ser acessadas da forma que ele precisa, isso sempre foi um questionamento do corpo clínico, que inclusive fez parte de pesquisas anteriores deste autor (CARVALHO, 2017).

Baseado neste cenário, essa pesquisa norteia-se pelo seguinte questionamento: *Como um chatbot pode ser utilizado para recuperar informações de um repositório Web, de forma que os seus metadados possam ser utilizados para resolver possíveis ambiguidades e imprecisões durante o processo de comunicação entre o usuário e o sistema?*

Além disso, outras questões secundárias podem contribuir para especificar o problema dessa pesquisa, como:

- Como aperfeiçoar os *chatbots* para serem utilizados em sistemas de recuperação de informação?
- A Ciência da Informação pode contribuir no processo de construção dos *chatbots* e de seus fluxos conversacionais para serem utilizados em sistemas de recuperação de informação?
- Os metadados podem contribuir na criação dos fluxos conversacionais dos *chatbots* de sistemas de recuperação de informação?

1.2 Hipóteses

A partir do cenário descrito a respeito do problema desta pesquisa, algumas hipóteses foram identificadas e que são descritas a seguir:

- Os metadados disponibilizados por diversos portais podem ser utilizados para construção de fluxos comunicacionais de agentes interativos.
- Durante uma conversa por meio de um *chatbot*, o número de respostas é inversamente proporcional à qualidade destas. Assim, a utilização da estrutura

dos metadados pode fornecer as ligações entre os termos e determinar sua relevância no diálogo.

- Um repositório disponibilizado na Internet que possui seu acervo descrito, de forma estruturada, por meio de metadados pode facilitar a construção de agentes comunicacionais.
- Durante um diálogo, a utilização de metadados no processo, possibilita o reconhecimento de novos termos a serem apresentados aos usuários, podendo facilitar o processo de recuperação de informações.

1.3 Tese

Considerando a recuperação de informação como um processo comunicacional e dialógico, problemas inerentes à comunicação, tais como subjetividade, imprecisão e ambiguidade, podem ser minimizados quando os metadados que descrevem e representam os documentos em um repositório podem ser utilizados no desenvolvimento de agentes conversacionais.

1.4 Proposta da pesquisa

A proposta deste trabalho aventa a criação de um modelo conceitual de recuperação de informação por meio de um *chatbot*, cuja etapa de construção de diálogos seja enriquecida por termos provindos dos metadados da base de dados do repositório.

Dessa forma, analisando a convergência entre as áreas da Ciência da Informação, dos agentes conversacionais com processamento de linguagem natural, sistemas de recuperação de informação e os metadados, foi idealizada essa tese como uma contribuição sobre esse tema e um aprofundamento para auxiliar a criação de sistemas mais simples de serem utilizados e mesmo assim ainda eficientes na busca de informações, cumprindo um grande papel social da inclusão digital por intermédio dos mais diferentes meios de acesso e permitindo aos desenvolvedores implementar novas interfaces para acesso e disponibilização das informações.

1.5 Objetivos

Os objetivos que orientam essa pesquisa foram divididos em dois tópicos, o objetivo geral e os objetivos específicos, que são apresentados em seguida.

1.5.1 Objetivo geral

O objetivo geral desta pesquisa é modelar conceitualmente um agente conversacional, como um *chatbot*, considerando os metadados que representam os documentos do repositório enquanto desempenha a função de um sistema de recuperação de informação.

Ao enriquecer os termos fornecidos pelo usuário durante o diálogo com informações provindas dos metadados, é possível aumentar a relevância dos resultados, e devido às características dos sistemas de bate-papo, linguagem natural em dispositivos móveis, comumente usados, avaliar se este tipo de comunicação poderia ser aplicado a qualquer repositório, ou até mesmo, se é possível disponibilizar qualquer repositório até as redes sociais e aplicativos de mensagens instantâneas.

1.5.2 Objetivos específicos

E como objetivos específicos desta pesquisa:

- Identificar, com base na literatura, além de informações oriundas da área da Ciência da Informação, o atual cenário da pesquisa em interfaces e modelos de agentes conversacionais e suas aplicações.
- Verificar a aplicabilidade de interfaces em linguagem natural em diferentes sistemas e em especial, sistemas de recuperação de informação.
- Analisar as etapas de pré-processamento, processamento e pós-processamento dos módulos envolvidos, dos dados e seus resultados.
- Verificar a eficiência e o desempenho da metodologia sugerida em um sistema de recuperação, de forma que durante o processo de comunicação o acesso aos metadados possam complementar os termos provindos do usuário.
- Elaborar um protótipo de agente conversacional com uma demonstração da viabilidade e usabilidade desta proposta e a verificação do seu desempenho para comprovação desta tese sobre um ambiente de teste desenvolvido e disponibilizado na Web especificamente para este trabalho, denominado ENANCIB WEB.

1.6 Justificativa

O estudo justifica-se pela necessidade de acesso a informações disponíveis em sistemas de armazenamento de dados, e com o advento de novas tecnologias ocorre a busca da aplicação de novas ideias com o intuito de aprimorar esse processo. Pesquisas acadêmicas são aplicadas a sistemas de recuperação de informação e na forma de interação entre as pessoas e as máquinas, além disso, sistemas de inteligência artificial nos trazem novas formas naturais de interação com os computadores.

Minha proposta é aplicar tais tecnologias, interfaces inovadoras, recuperação de informação em repositórios de documentos na Internet, na melhoria da escolha dos termos utilizados na busca de documentos. Além disso, há muitos fatores externos que podem indicar uma necessidade de pesquisas desta linha, como o surgimento de novos meios de acesso por meio de novas tecnologias, a presença cada vez maior das pessoas conectadas na Internet o tempo todo, a inclusão social permitida por sistemas desse tipo, a presença de muitas pessoas já interagindo nas redes sociais e nos aplicativos de comunicação e a novidade de pesquisas dessa linha, argumentos que serão demonstrados a seguir.

Para evidenciar as mudanças que vem ocorrendo nos últimos tempos no Brasil, CETIC.BR (2022) elabora anualmente, desde 2005, relatórios sobre a presença de tecnologias nos lares dos brasileiros e seus usos. Em 2022, foram entrevistados 23.950 domicílios em todo o território nacional. Foi utilizada a metodologia face a face entre outubro de 2021 e março de 2022. O número de domicílios brasileiros em 2021 com acesso à Internet chegou a 82% do total, um acréscimo de 11% de domicílios comparado a 2019 (CETIC.BR, 2020a). Pela terceira vez, mais da metade dos domicílios das classes D e E estavam conectados à Internet, passou de 30% em 2015 para 61% em 2021. Outro fato importante, evidenciou que o telefone celular foi o principal dispositivo de acesso à Internet com 99% de usuários com acesso entre 2019 e 2021. O acesso por meio de computadores continua caindo a cada ano, chegando a 36% em 2021.

Os dados mostram que a Internet passou a fazer parte das vidas das pessoas em qualquer lugar que estiverem no Brasil. Até mesmo nas áreas rurais, hoje, o acesso à Internet é possível e com qualidade satisfatória, onde a pesquisa mostrou que 71% da população de áreas rurais já tem acesso à Internet, sendo a taxa de 83% na área urbana.

Sobre os principais dispositivos para acessar a Internet, o celular é o preferido entre 99% dos lares, chegando a 58% dos usuários como único meio de acesso, e nas classes mais pobres

(D e E) esse valor chega a 90% em 2020. As taxas de uso exclusivo pelo celular também podem ser diferentes entre a população preta (65%), parda (69%) e branca (54%).

Também demonstrou que o acesso por meio de um computador de mesa ou portátil vem mudando bastante, em 2014 o acesso era feito por 80% da população por meio desse dispositivo, essa taxa foi caindo até chegar a 36% em 2021, e se for aferido somente os computadores de mesa, esse valor despencou de 54%, em 2014, para 23% em 2019. Já sobre o acesso por meio de uma televisão, esse valor subiu vinte pontos em relação a 2018 chegando a 50% no ano de 2021, o que mostra uma mudança bastante interessante dos meios de acesso à Internet. O relatório de 2020, mostrou que o número de usuários de Internet pela televisão ultrapassou o número de usuários por computadores, de 44% a 42%.

Ainda para justificar uma pesquisa nessa área, podemos verificar a existência de outros trabalhos finalizados ou em desenvolvimento no país, vide Quadro 1. Na busca da compreensão do cenário atual desse campo no Brasil e nas mais diversas áreas de pesquisas, foram feitas buscas na BDTD, como forma de verificar o que havia a respeito desses assuntos, e foram poucos os trabalhos relevantes encontrados, e praticamente nada sobre o escopo desse projeto.

Foram escolhidas várias palavras-chaves relacionadas ao assunto e utilizadas no sistema de busca avançada⁴ da plataforma, além de não selecionar mais nenhuma outra informação, de forma que pudesse ser o mais abrangente e pudesse trazer ao conhecimento toda e qualquer pesquisa desenvolvida.

No Quadro 1, estão demonstradas todas as buscas efetuadas na data de 30/09/2021, bem como as combinações de palavras usadas para formular as pesquisas. Todos os resultados positivos tiveram seus resumos analisados e assim foi determinado se continham relevância para serem selecionados, e então nesse momento, foram lidos os trabalhos completos para constatação de sua importância para o desenvolvimento deste trabalho. Como pode ser observado, quando se busca uma convergência entre as áreas de *chatbots*, recuperação de informação e metadados, existem poucas pesquisas que resultam dessa intersecção, o que justifica a investigação desse trabalho na busca da melhoria da qualidade dos resultados, de acordo com as particularidades de sistemas de interfaces distintos de meios de acesso tais como celulares, televisores ou outros dispositivos que possuam tal função.

⁴ Disponível em: <https://bdtb.ibict.br/vufind/Search/Advanced>. Acesso em: 21 ago. 2021.

Quadro 1 – Resultado das pesquisas na BDTD

Termos	Resultados Encontrados	Títulos dos trabalhos
<i>chatbot</i>	23 resultados (0 relevantes)	
<i>chatbot</i> E "recuperação de informação"	0 resultados	
<i>chatbot</i> E "recuperação da informação"	0 resultados	
<i>chatbot</i> E " <i>information retrieval</i> "	0 resultados	
"assistente virtual"	8 resultados (0 relevantes)	
"assistentes virtuais"	5 resultados (0 relevantes)	
<i>chatterbot</i>	19 resultados (5 relevantes)	1 - Um <i>Chatterbot</i> para Criação e Desenvolvimento de Ontologias com Lógica de Descrição ⁵ 2 - <i>Chatterbot</i> para criação e refinamento de ontologias em lógica de descrições ⁶ 3 - iAIML: um mecanismo para o tratamento de intenção em <i>Chatterbots</i> ⁷ 4 - Avaliação de <i>Faqbots</i> através da ferramenta <i>Autochatter</i> ⁸ 5 - <i>A study of the use of natural language processing for conversational agents</i> ⁹
"agentes inteligentes"	152 resultados (4 relevantes)	1 - SemanticAgent, uma plataforma para desenvolvimento de agentes inteligentes ¹⁰ 2 - Arquitetura para recuperação de objetos de aprendizagem – uma abordagem baseada em agentes inteligentes e <i>relevance feedback</i> ¹¹ 3 - Agentes inteligentes artificiais ¹² 4 - Arquitetura de software baseada em agentes para gerenciamento de portfólio de fontes de informação existentes na web ¹³
<i>chatbot</i> E metadado	0 resultados	
<i>chatterbot</i> E metadado	0 resultados	
<i>chatbot</i> E <i>metadata</i>	0 resultados	
<i>chatterbot</i> E <i>metadata</i>	0 resultados	

Fonte: Autor (2021)

Além das buscas na BDTD para avaliar a atual situação das pesquisas sobre este assunto no Brasil, foram feitas buscas na *Networked Digital Library of Theses and Dissertations* ¹⁴(NDLTD). É uma organização internacional que se dedica a preservar e disseminar teses e dissertações eletrônicas de todo o mundo. Em 12 de outubro de 2022, de acordo com a própria

⁵ Disponível em: https://bdtd.ibict.br/vufind/Record/UFPE_fae116ff5999600a7ee232254e99b200. Acesso em: 30 set. 2021.

⁶ Disponível em: https://bdtd.ibict.br/vufind/Record/UFPE_c19ba8bfac3f7408c63fb22f17f3b522. Acesso em: 30 set. 2021.

⁷ Disponível em: https://bdtd.ibict.br/vufind/Record/UFPE_3343f7bf1908834f0d786c5ff11e5464. Acesso em: 30 set. 2021.

⁸ Disponível em: https://bdtd.ibict.br/vufind/Record/UFSC_c6dda4175e81dd960c8e288fff9374d5. Acesso em: 30 set. 2021.

⁹ Disponível em: https://bdtd.ibict.br/vufind/Record/URGS_47dabc35acff51b3b2a3e47b39f0c7e8. Acesso em: 30 set. 2021.

¹⁰ Disponível em: https://bdtd.ibict.br/vufind/Record/USP_0486615f74d74238c965915a6ec08f6b. Acesso em: 30 set. 2021.

¹¹ Disponível em: https://bdtd.ibict.br/vufind/Record/UNSP_fdd701d54813f8416235e7b0039f907c. Acesso em: 30 set. 2021.

¹² Disponível em: https://bdtd.ibict.br/vufind/Record/PUC_SP-1_c87b893362209b04e8626df0b098254f. Acesso em: 30 set. 2021.

¹³ Disponível em: https://bdtd.ibict.br/vufind/Record/UFSC_5ecb9e0db5a7cbe96fc9ca9f2b714b27. Acesso em: 30 set. 2021.

¹⁴ Disponível em: <https://ndltd.org>. Acesso em: 12 out. 2022.

página na Web, continham em seu arquivo 6.346.399 teses eletrônicas e dissertações, todas disponíveis para consulta por sua ferramenta de busca¹⁵.

Foram efetuadas buscas utilizando os mesmos termos aplicados à BDTD, mas traduzidos para o inglês, e a metodologia de análise dos trabalhos foi a mesma na busca de trabalhos relevantes, e principalmente na mesma linha desta pesquisa. Assim como no Brasil, por meio da BDTD, foram poucos trabalhos relevantes recuperados por meio dos termos escolhidos e nenhum trabalho com o escopo deste trabalho, o que pode ser verificado no Quadro 2, os títulos e os *links* para acesso aos mesmos.

Quadro 2 – Resultado das pesquisas na NDLTD

Termos	Resultados Encontrados	Títulos dos trabalhos relevantes
<i>chatbot</i>	366 resultados	<i>Utterances classifier for chatbots intents</i> ¹⁶
<i>chatbot and information and retrieval</i>	13 resultados	<i>A data-driven approach for a chatbot using transcripts from a TV-series</i> ¹⁷
<i>virtual and assistant</i>	2683 resultados	<i>Diseño de un servicio de respuesta automático sobre el gasto público de Chile mediante un asistente virtual de interfaz conversacional</i> ¹⁸
<i>virtual and assistant and metadata</i>	9 resultados	
<i>chatterbot</i>	33 resultados	
<i>intelligent and agent</i>	3336 resultados	<i>Making Chatbots More Conversational: Using Follow-Up Questions for Maximizing the Informational Value in Evaluation Responses</i> ¹⁹
<i>intelligent and agent and metadata</i>	31 resultados	
<i>chatbot and metadata</i>	1 resultado	<i>Design of a FAQ Chatbot System Based on Google Dialogflow</i> ²⁰
<i>chatterbot and metadata</i>	0 resultados	<i>Dialogue Systems Using Web-based Language Tools</i> ²¹
		<i>Chatbot for Information Retrieval from Unstructured Natural Language Documents</i> ²²
		<i>Response Quality in Human-chatbot Collaborative Systems</i> ²³
		<i>Chatbot for Information Retrieval from Unstructured Natural Language Documents</i> ²⁴
		<i>Conversational Chatbots with Memory-based Question and Answer Generation</i> ²⁵
		<i>Development of Retrieval-based Chatbot - Using PTT Gossip Articles as a Knowledge Base</i> ²⁶

15 Disponível em: <http://search.ndltd.org>. Acesso em: 12 out. 2022.

16 Disponível em: <http://urn.kb.se/resolve?urn=urn:nbn:se:kth:diva-233362>. Acesso em: 12 out. 2022.

17 Disponível em: <http://urn.kb.se/resolve?urn=urn:nbn:se:kth:diva-157635>. Acesso em: 12 out. 2022.

18 Disponível em: <http://repositorio.uchile.cl/handle/2250/168100>. Acesso em: 12 out. 2022.

19 Disponível em: <http://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-393269>. Acesso em: 12 out. 2022.

20 Disponível em: <http://ndltd.ncl.edu.tw/handle/u735hg>. Acesso em: 12 out. 2022.

21 Disponível em: <http://urn.kb.se/resolve?urn=urn:nbn:se:umu:diva-140733>. Acesso em: 12 out. 2022.

22 Disponível em: <http://urn.kb.se/resolve?urn=urn:nbn:se:hh:diva-40532>. Acesso em: 12 out. 2022.

23 Disponível em: <http://hdl.handle.net/10919/106683>. Acesso em: 12 out. 2022.

24 Disponível em: <http://urn.kb.se/resolve?urn=urn:nbn:se:hh:diva-40532>. Acesso em: 12 out. 2022.

25 Disponível em: <http://urn.kb.se/resolve?urn=urn:nbn:se:liu:diva-171927>. Acesso em: 12 out. 2022.

26 Disponível em: <http://ndltd.ncl.edu.tw/handle/8h7e9c>. Acesso em: 12 out. 2022.

		<i>A Question-Answering Conversational Agent with Recommendations Based on a Domain Ontology</i> ²⁷ <i>Chatbot for Information Retrieval from Unstructured Natural Language Documents</i> ²⁸
--	--	---

Fonte: Autor (2022)

Outra justificativa que pode ser atribuída a esta pesquisa é o grau de novidade, segundo AHMAD (2021), o processamento de linguagem natural (PLN) chegou a tal nível que pode contribuir de maneira significativa aos sistemas de recuperação de informação, pois a característica principal desses sistemas é aplicar as redes neurais para entender a linguagem do usuário em um nível mais profundo do que uma pesquisa efetuada por meio de palavras-chave. Isso permite que as buscas extrapolem uma variedade maior de conteúdo relevante, ao mesmo tempo em que mostram resultados com maior precisão. A utilização de sistemas de recuperação de informação neurais e a busca semântica só começaram a ser apresentadas ao público em 2017, e ainda complementa, o primeiro exemplo comercial desse sistema foi o Amazon Kendra²⁹, lançado no início de 2020, enquanto o *Azure Cognitive Search*³⁰, lançado em abril de 2021, e *ZIR Semantic Search*³¹ são mais recentes.

Outro fato aliado a isso são as pesquisas e investimentos com foco em temas específicos da área procurando novos modelos de aprendizado de máquina, robótica cognitiva, lógica probabilística e ontologia computacional (USP, 2018), além de um maior interesse de empresas de diversos setores, com o desejo de enriquecer seus canais digitais com aplicações que combinam computação cognitiva e interfaces mais amigáveis, baseadas em fala, voz e assistentes virtuais (IBM, 2018).

1.7 Procedimentos Metodológicos

A natureza aplicada desse trabalho segue os preceitos de uma pesquisa exploratória que conforme Gil (2002, p. 41) visa proporcionar maior familiaridade com o problema, de modo a torná-lo mais explícito ou construir hipóteses, pode-se utilizar da pesquisa bibliográfica em fontes primárias de informação como livros, artigos, teses, dissertações, monografias, entre outros referentes ao assunto, para verificar os conceitos utilizados neste trabalho e as metodologias disponíveis para a realização do mesmo. Assim, realizando uma pesquisa a fim de identificar o “estado da arte” das pesquisas a respeito da aplicação das interfaces cognitivas,

27 Disponível em: <http://doi.org/10.17771/PUCRio.acad.50180>. Acesso em: 12 out. 2022.

28 Disponível em: <http://urn.kb.se/resolve?urn=urn:nbn:se:hh:diva-40532>. Acesso em: 12 out. 2022.

29 Disponível em: <https://aws.amazon.com/pt/kendra>. Acesso em: 21 ago. 2021.

30 Disponível em: <https://docs.microsoft.com/pt-br/azure/search/search-what-is-azure-search>. Acesso em: 21 ago. 2021.

31 Disponível em: <https://zir-ai.com>. Acesso em: 21 ago. 2021.

e ainda mais especificamente, na forma de *chatbots* em linguagem natural no acesso a informações em sistemas de armazenamento de dados.

O ambiente determinado para essa pesquisa é constituído a partir dos alicerces da recuperação de informação como ferramenta para permitir o acesso a informações persistidas em banco de dados de forma estruturada, com os seus metadados qualificados, e disponíveis para acesso por meio de uma interface conversacional presente em uma rede social ou aplicativo de mensagens instantâneas. O estudo se baseia em como a presença dos metadados dos documentos acessados estando disponíveis no processo de construção dos diálogos podem influenciar na qualidade dos resultados, apresentados ao usuário durante a conversa, na busca das informações relevantes. Como cenário para este trabalho, foi desenvolvido um repositório de dados com trabalhos apresentados em todas as edições dos Encontros Nacionais de Pesquisa e Pós-Graduação em Ciência da Informação (ENANCIB), minerados na Internet e compilados juntamente com seus metadados, de 1994 a 2021, para que possam servir de banco de dados para a exemplificação deste modelo proposto.

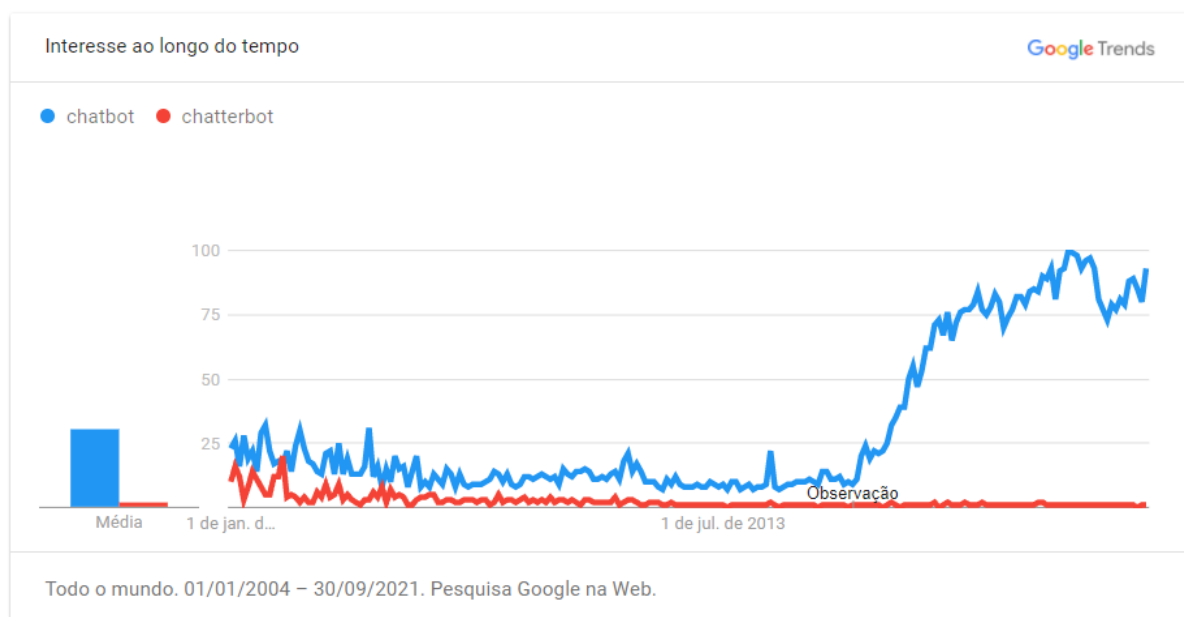
1.8 Terminologia Adotada

Foi escolhido o termo *chatbot* para ser utilizado neste trabalho devido à sua evidência mostrada pelo sistema de buscas Google. De acordo com o Google Trends (2021), os números no gráfico

representam o interesse de pesquisa relativo ao ponto mais alto no gráfico de uma determinada região em um dado período. Um valor de 100 representa o pico de popularidade de um termo. Um valor de 50 significa que o termo teve metade da popularidade. Uma pontuação de 0 significa que não havia dados suficientes sobre o termo.

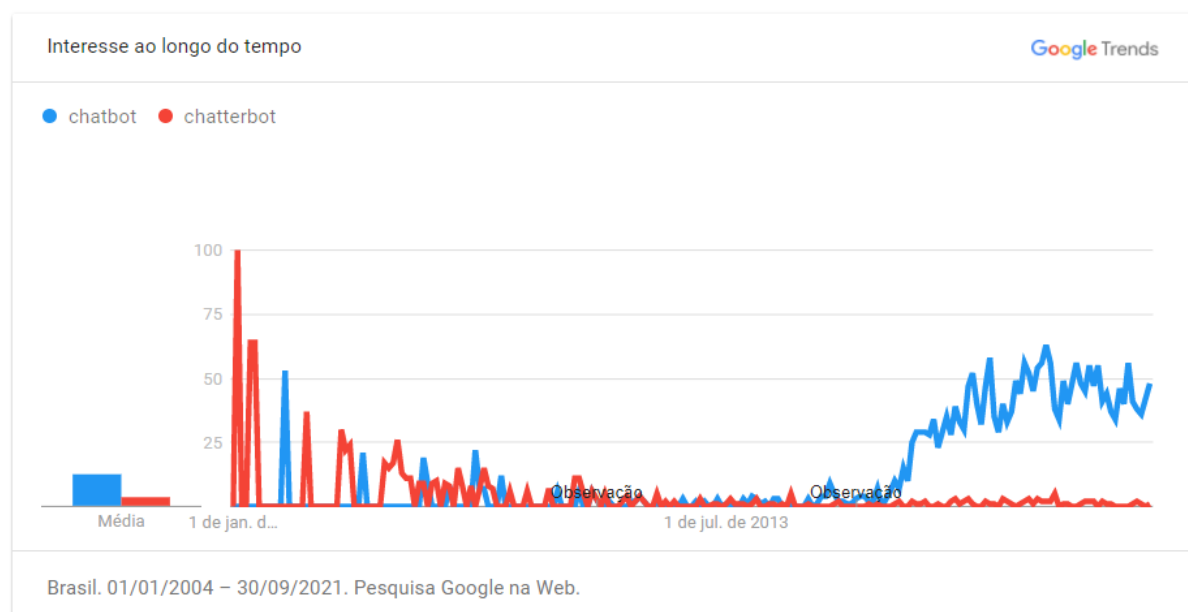
Ambos os termos *chatbot* e *chatterbot* são usados em diversas pesquisas e trabalhos pelo mundo, mas a plataforma mostra sua maior utilização destes termos para buscar textos relacionados a esse tema. Analisando os resultados do Google Trends (2021), verifica-se que o termo *chatbot* está mais associado a aplicativos de bate-papo e sistemas de desenvolvimento destes, já o termo *chatterbot* está mais associado a linguagens de programação de *chatbots*. Isso acontece com uma média maior sobre o termo *chatbot* em buscas no mundo todo, conforme Figura 1, e acontece um cenário bastante parecido no Brasil, conforme a Figura 2. Mais informações sobre o termo e suas origens serão discutidas no capítulo específico sobre esse tema.

Figura 1 – Pesquisas dos termos *chatbot* e *chatterbot* no Mundo



Fonte: Google Trends (2021)

Figura 2 – Pesquisas dos termos *chatbot* e *chatterbot* no Brasil



Fonte: Google Trends (2021)

Cabe ressaltar que, em 01 de janeiro de 2016, o Google alega ter feito uma alteração em seu algoritmo para coleta de dados mais eficiente, o que pode ser observado nas Figuras 1 e 2 um aumento na taxa de buscas do termo *chatbot*. Além disso, observou-se que a partir deste mesmo ano os sistemas de *chatbots* começaram a ser disponibilizados em grande parte pelos principais serviços de mensagem em tempo real do mundo, como Facebook Messenger, Skype, Slack, Telegram e WeChat (principal aplicativo de mensagem na China). Estas empresas

começaram a oferecer ferramentas para criação e disponibilização de *chatbots*, por meio de APIs (do Inglês *Application Programming Interface*, que traduzido é Interface de Programação de Aplicações), para integração aos sistemas corporativos. Em 2017, *chatbots* que utilizavam inteligência artificial para interagir com os humanos em forma de linguagem natural começaram a ser apresentados e a atrair o interesse das empresas e ser utilizado pelos consumidores (CIO, 2018).

Uma vasta gama de termos para se referenciar aos componentes destes sistemas podem ser encontrados. Os fabricantes de soluções (Amazon, Microsoft, IBM, Google) para o desenvolvimento de *chatbots* costumam denominar as ferramentas como pertencentes de sistemas de computação cognitiva. Já os usuários e pesquisadores se referem a tal como *chatbots*, *chatterbots*, agentes inteligentes, assistentes virtuais, agentes cognitivos, interfaces cognitivas, interfaces inteligentes, entre tantas outras. Para este trabalho foi determinado que um *software* na forma de um agente de bate-papo deverá ser tratado como um *chatbot*, e no caso de plataformas com uma variedade de sistemas, como agentes conversacionais.

1.9 Organização do trabalho

Quanto à estrutura desse trabalho, ele foi constituído de 6 capítulos e foram distribuídos da seguinte forma:

Capítulo 1 – Apresenta a introdução ao trabalho com a apresentação do cenário atual das tecnologias e processos envolvidos e a descrição das partes do trabalho como identificação do tema, uma caracterização do problema, as hipóteses levantadas, a proposição e a tese, o objetivo geral e específicos, justificativa e os procedimentos metodológicos adotados para a execução do trabalho.

Capítulo 2 – Apresenta a área da Recuperação de Informação e sua origem, as definições, autores, entre outras informações, e o processo da recuperação de informação com suas etapas definidas a partir de um modelo, processamento de linguagem natural e suas particularidades e a recuperação de informação como um processo comunicacional.

Capítulo 3 – Demonstra como os *chatbots* podem ser utilizados, sua definição, características, tipos, diferenças entre eles e os assistentes digitais, plataformas conversacionais e os módulos que compõem um sistema conversacional.

Capítulo 4 – Serão apresentadas informações a respeito dos metadados e seus usos, a partir de exemplos, poderão ser visualizadas formas de os aplicar a sistemas disponíveis na Internet e que poderiam fornecer informações para outros sistemas, como o modelo proposto neste trabalho.

Capítulo 5 – Neste capítulo será demonstrado como a aplicação dos metadados no processo de construção de diálogos de um *chatbot* pode elevar o grau de sucesso ao recuperar mais documentos relevantes em menos interações com o usuário. O desenvolvimento do modelo conceitual e do repositório dos documentos utilizados para a prova do conceito.

Capítulo 6 – Aqui serão discutidos os resultados a respeito da aplicação do modelo e avaliação da sua eficácia com a demonstração dos dados avaliados e o resultado do projeto. Também são apresentadas algumas discussões à cerca dos resultados obtidos e seguido de reflexões, limitações e contribuições do trabalho terminando com sugestões para pesquisas futuras.

2. RECUPERAÇÃO DE INFORMAÇÃO

A criação do termo "recuperação de informação" é geralmente atribuída ao matemático e cientista da computação Calvin Northrup Mooers. Em artigo de 1951, Mooers definiu e apresentou os problemas a serem abordados por esta nova disciplina.

Recuperação de informação é o nome dado ao processo ou método pelo qual um usuário de informação é capaz de converter a sua necessidade de informação em uma lista de citações a documentos em um acervo contendo informações úteis para ele.

[...]

Recuperação de informação abrange os aspectos intelectuais da descrição da informação e sua especificação para a busca, e também quaisquer sistemas, técnicas ou máquinas que são utilizadas para realizar a operação.

[...]

O assunto de cada documento ou outra unidade de informação é caracterizado ou descrito por meio de um conjunto de "descritores" tirado de um vocabulário formal de tais termos. Uma "lista de cabeçalho de assuntos" remeterá a uma aproximação grosseira do seu significado. (MOOERS, 1951, p. 25, tradução nossa).³²

Na definição de Mooers é possível identificar três elementos básicos da recuperação de informação: o "usuário de informação" e sua necessidade de informação; os documentos caracterizados ou descritos "por meio de um conjunto de descritores"; e um elemento mediador: "quaisquer sistemas, técnicas ou máquinas".

³² Information retrieval is the name for the process or method whereby a prospective user of information is able to convert his need for information into an actual list of citations to documents in storage containing information useful to him.

[...]

Information retrieval embraces the intellectual aspects of the description of information and its specification for search, and also whatever systems, techniques, or machines that are employed to carry out the operation.

[...]

The subject matter of each document or other unit of information is characterized or described by means of a set of "descriptors" taken from a formal vocabulary of such terms. A "subject heading list" will call to mind a rough approximation of what is meant here.

Anteriormente, no ano de 1950, em um boletim técnico publicado em um encontro da *Association of Computing Machinery* (ACM), Mooers enfatizava o caráter linguístico (não numérico) do processo de recuperação de informação:

[...] A recuperação da informação é um problema não numérico em parte porque a maior parte da comunicação humana é verbal, mas mais importante porque a maioria das ideias ou conceitos não podem ser mapeados em um espaço Euclidiano de 3 ou mais dimensões. Embora haja valores de escala para a representação de algumas informações, estas são relativamente poucas e sem importância. Conceitos espaciais e métricos não se aplicam à maioria das informações, pelo menos não nos níveis mais simples. No entanto, embora o problema de recuperação de informação não seja numérico, não parece haver alternativa ao uso de técnicas digitais para sua solução. Sistemas de recuperação de informação digital que empregam máquinas já estão operando, e seu grau de sucesso parece indicar que esta é a direção do progresso. (Mooers, 1950, p.3) ³³

Em consonância com Mooers, FERNEDA (2013) apresenta os principais elementos envolvidos no processo de recuperação de informação, destacando a característica linguística de tais elementos:

Recuperar informação envolve, por um lado, um acervo documental que deve ser representado por expressões linguísticas que resumem seu conteúdo informacional. Por outro lado, temos seres humanos que tentam descrever linguisticamente as suas necessidades de informação a fim de obterem documentos relevantes para satisfazer tais necessidades. Portanto, um sistema de recuperação de informação é um ambiente linguístico mediador na comunicação entre um estoque de informação e seus requisitantes. (FERNEDA, 2013, p. 6)

O objetivo de um sistema de recuperação de informação (SRI) é estimar a relevância de cada item de informação em relação à necessidade de informação de um determinado usuário. Essa tarefa é complexa, pois é permeada de subjetividade, imprecisão e incerteza. Assim, A subjetividade é uma propriedade intrínseca a qualquer sistema de recuperação, pois está relacionado ao próprio conceito de relevância. Em resposta à uma mesma busca, um determinado documento pode ser considerado totalmente relevante para um usuário e não ser relevante para outro. Somente o usuário é capaz de julgar se um documento é relevante para a sua necessidade (CRESTANI; PASI, 2003).

³³ Information retrieval is a non-numerical problem in part because the most of human communication is verbal, but more important because most ideas or concepts cannot be mapped into a Euclidean 3-space, or higher space. While there are scale readings for the representation of some information, these are relatively few and unimportant. Spatial and metrical concepts do not apply to most information, at least not at the simpler levels. Yet, though the information retrieval problem is non-numerical, there does not seem to be any alternative to the use of digital techniques for its solution. Digital information retrieval systems employing machines are already operating, and their degree of success seems to indicate that this is the direction of progress.

Além do conceito de relevância, a ambiguidade e a imprecisão também afetam o processo de recuperação de informação. Esses aspectos estão presentes principalmente na tarefa de traduzir uma determinada necessidade de informação em uma expressão de busca que melhor a represente, mas também podem afetar no processo de representação da informação (MARTINES; OLIVEIRA; SABBAG, 2019; CARVALHO, 1995). Pesquisas demonstram que o usuário muitas vezes não tem uma imagem clara do que está procurando e representa sua necessidade utilizando termos vagos, imprecisos ou mesmo ambíguos (INGWERSEN, 1992). Para contribuir com este processo, alguns recursos interativos podem ser implementados a fim de auxiliar os usuários na obtenção de resultados mais relevantes para sua necessidade de informação. Também é possível disponibilizar recursos para auxiliar o usuário no julgamento de relevância dos documentos resultantes de uma determinada busca.

A maioria dos sistemas de recuperação de informação oferece uma abordagem relativamente simples do conceito de relevância e da subjetividade e imprecisão inerentes a esse processo. Alguns sistemas mais avançados tentam lidar com a subjetividade, a incerteza e a imprecisão propondo modelos mais complexos e sofisticados.

2.1 Recuperação de Informação como Área de Pesquisa

Segundo Popper (1972, apud SARACEVIC, 1996, p. 41) “... não somos estudantes de algum assunto, mas estudantes de problemas. E os problemas podem cruzar a fronteira de qualquer assunto ou disciplina”. Assim, pode-se posicionar a recuperação de informação como um problema que pode ser abordado por diversas áreas científicas.

No campo da Ciência da informação, as pesquisas em recuperação de informação se concentram prioritariamente no comportamento humano no contexto da busca, recuperação e uso da informação (WILSON, 2000). Na Ciência da Computação, a recuperação de informação se refere à criação e o aperfeiçoamento de algoritmos e sistemas relacionados à representação, organização e seleção de informações relevantes que atendem às necessidades do usuário a partir da grande quantidade de informações não estruturadas armazenadas no computador (MARCHIONINI, 1997).

Na Ciência da Computação a recuperação de informação se firmou como uma subárea de pesquisa, com um conjunto de pesquisadores especializados, periódicos, congressos e instituições acadêmicas direcionadas para essa área. Na Ciência da Informação não existe essa

especificidade, e a recuperação se refere ao objetivo intrínseco do "processamento técnico" (catalogação, indexação, classificação).

Segundo Saracevic (1995), na Ciência da Informação a maioria dos esforços e recursos são dedicados a algum problema relacionado ao processo de recuperação da informação, sendo essa a sua maior fonte de suas relações interdisciplinares. Saracevic (1996, p.47) define a Ciência da Informação como:

[...] um campo dedicado às questões científicas e à prática profissional voltadas para os problemas da efetiva comunicação do conhecimento e de seus registros entre os seres humanos, no contexto social, institucional ou individual do uso e das necessidades de informação. No tratamento destas questões são consideradas de particular interesse as vantagens das modernas tecnologias informacionais.

Os problemas relacionados à disseminação do conhecimento, como a recuperação de informação, não podem ser resolvidos no âmbito de uma única disciplina, um único campo científico. Assim, a interdisciplinaridade da Ciência da Informação é imposta pela variedade de especialidades acadêmicas e profissionais que se ocuparam com os problemas abordados por essa área. Entre os pioneiros da Ciência da Informação havia engenheiros, bibliotecários, químicos, linguistas, filósofos, psicólogos, matemáticos, cientistas da computação, entre outros profissionais vindos de diferentes campos científicos (SARACEVIC, 1996).

A relação entre a Ciência da Informação com a Ciência da Computação acontece principalmente pelo interesse comum na busca de soluções de problemas relacionados à recuperação de informação. Denning *et al* (1989) define a Ciência da Computação com:

A disciplina da computação é o estudo sistemático dos processos algorítmicos que descrevem e transformam informações; sua teoria, análise, projeto, eficiência, implementação e aplicação. A questão fundamental subjacente a toda computação é "O que pode ser (eficientemente) automatizado?"³⁴

A Ciência da Computação, portanto, trata de processos algorítmicos que descrevem e transformam informações, enquanto a Ciência da Informação trata da natureza da informação e sua comunicação para uso por humanos. Ambas as Ciências estão interrelacionadas e são complementares, na medida em que buscam soluções para a organização, a representação e a disseminação de seu objeto comum: a informação.

³⁴ The discipline of computing is the systematic study of algorithmic processes that describe and transform information: their theory, analysis, design, efficiency, implementation, and application. The fundamental question underlying all of computing is, "What can be (efficiently) automated?"

Em 1945, após o fim da Segunda Grande Guerra e com o surgimento dos primeiros computadores, pesquisadores como o engenheiro Vannevar Bush (BUSH, 1945) manifestava em seu artigo "*As We May Thing*" a importância e a urgente necessidade da organização do conhecimento científico. Segundo Johnston e Webber (2005, p.109):

O artigo de Bush pode ser considerado como uma descrição de um microcosmo da sociedade da informação, com os limites fortemente traçados pelos interesses e experiências de um grande cientista da época, em vez dos espaços de conhecimento mais abertos do século XXI. [...]

Bush fornece uma visão central da importância da informação para a sociedade industrial/científica, usando a imagem de uma "explosão da informação" decorrente das demandas inéditas de produção científica e aplicação tecnológica da Segunda Guerra Mundial. Ele descreve uma versão da ciência da informação como uma disciplina-chave dentro da prática dos domínios do conhecimento científico e técnico. Sua visão abrange os problemas de sobrecarga de informações e a necessidade de criar mecanismos eficientes para controlar e canalizar as informações para uso.³⁵

Neste mesmo ano, Friedrich A. Hayek (1945), publicou artigo intitulado "*The Use of knowledge in society*", em que discute que tomadores de decisões precisam ter acesso a informações para definir as ações necessárias. Calvin N. Mooers (1950, 1951) apresenta o conceito de "recuperação de informação" e dá um rótulo, um nome a uma área de pesquisa que abarcava problemas e preocupações de diversos pesquisadores da época.

Nas décadas que se seguiram, com a evolução dos computadores e a Ciência da Computação, diversos estudos começam a aparecer, principalmente em relação a práticas automatizadas de tratamento de informação. As experiências de Hans Peter Luhn, em 1958, trouxeram ideias e tecnologias que ainda hoje são utilizadas (SCHULTZ, 1968).

Gerard Salton, matemático e cientista da computação, foi talvez o principal pesquisador no campo da Recuperação da Informação. Considerado como o pai da Recuperação de Informação, é mais conhecido por desenvolver o Modelo Espaço Vetorial, implementado no sistema SMART. Foi responsável em propor conceitos e técnicas que ainda hoje são utilizadas em sistemas de recuperação e nos mecanismos de busca Web (SALTON; WONG; YANG,

³⁵ Bush's paper might be regarded as describing a microcosm of the information society, with the boundaries tightly drawn by the interests and experiences of a major scientist of the time, rather than the more open knowledge spaces of the 21st century. He was looking forward speculatively to where we now are and we will look back through his speculations, as we reflect on the current situation.

Bush provides a core vision of the importance of information to industrial/scientific society, using the image of an "information explosion" arising from the unprecedented demands on scientific production and technological application of World War II. He outlines a version of information science as a key discipline within the practice of scientific and technical knowledge domains. His view encompasses the problems of information overload and the need to devise efficient mechanisms to control and channel information for use.

1975). Ao final de sua carreira, Salton desenvolveu importantes pesquisas na área de análise e sumarização automática de texto. No decorrer de sua vida como professor e pesquisador publicou mais de 150 artigos e 5 livros.

Na conclusão de seu artigo sobre a história da Recuperação de Informação, Sanderson e Croft (2012) argumentam que:

O século 20 e o início do século 21 foram transformadores na forma como as pessoas acessavam as informações. Em 1912, uma pessoa com necessidade de informação provavelmente iria a uma biblioteca local e, usando um catálogo de fichas, localizaria livros ou documentos que atendessem a essa necessidade. Devido à relativa inconveniência de acessar informações dessa maneira, essa pessoa provavelmente procuraria apenas responder a um pequeno número de perguntas. O escopo da informação disponível para as pessoas seria limitado pelo tamanho de sua biblioteca; para um pequeno número de necessidades muito importantes, um empréstimo entre bibliotecas poderia ser providenciado. Por causa da ubiquidade da busca baseada na Web, não é preciso dizer qual é o estado da arte atual: para aqueles com uma conexão à Internet, pode-se acessar instantaneamente centenas de *terabytes* de páginas, vídeos, notícias, imagens, redes sociais, mídia, livros digitalizados, trabalhos acadêmicos, música, programas de televisão e filmes; quase sempre através de motores de busca. Nos últimos anos, o acesso também tem sido possível a partir de um smartphone. A única coisa em comum entre a situação de hoje e a de 100 anos atrás é que ambos os serviços são geralmente gratuitos.³⁶

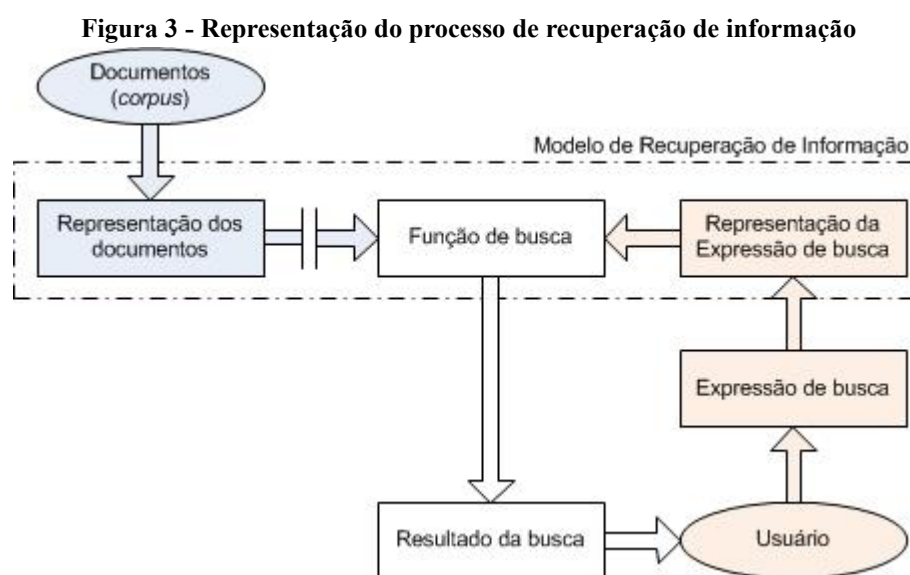
Em vista da evolução tecnológica observada nos últimos anos, pode-se considerar possíveis direções para o futuro da Recuperação de Informação. Com a crescente quantidade de informações disponíveis em meios eletrônicos, a importância dos meios de recuperação de informação crescerá em proporções semelhantes. Porém, considerando o avanço tecnológico observado desde as primeiras pesquisas em Recuperação de Informação até o surgimento da Web, a busca por informações relevantes e úteis é ainda uma tarefa muitas vezes árdua e frustrante. Ferneda (2013, p.11) argumenta que aparentemente os principais elementos relacionados ao processo de recuperação de informação são alheios aos avanços tecnológicos ou, pelo menos às soluções tecnológicas disponíveis atualmente.

³⁶ The 20th and early 21st centuries were transformational in the way people accessed information. In 1912, a person with an information need would probably go to a local library and, using a card catalog, locate books or documents that hopefully answered that need. Because of the relative inconvenience of accessing information in that way, that person would most likely only seek to answer a small number of questions. The scope of information available to people would be limited by the size of their library; for a small number of very important needs, a loan across libraries might have been arranged. Because of the ubiquity of web-based search, it need hardly be said what the current state of the art is: for those with an Internet connection, one can instantaneously access hundreds of terabytes of web pages, video clips, news, images, social media, scanned books, academic papers, music, television programs, and films; almost always through search engines. In the last few years, the access has been also possible from a mobile phone. Just about the only thing in common between the situation today and 100 years ago is that both services are generally free at the point of use.

2.2 O Processo de Recuperação de Informação

O processo de recuperação de informação envolve, por um lado, um acervo de documentos, geralmente representados por termos de indexação cujos significados remetem aos assuntos tratados por eles. Por outro lado, as pessoas traduzem as suas necessidades de informação por meio de um conjunto de termos a fim de encontrar documentos que venham a satisfazer tais necessidades. Um sistema de recuperação de informação é um elemento mediador entre um acervo de documentos e os usuários que buscam por documentos.

Segundo Fernalda (2012), o processo de recuperação de informação busca identificar, dentre um conjunto de documentos, aqueles que atendam a uma determinada necessidade de informação. Na Figura 3 é apresentada uma representação do processo de recuperação de informação.



Fonte: adaptado de Fernalda (2012, p.14)

2.2.1 Documentos (corpus)

O Aulete Digital ³⁷ define do verbete "documento" como:

(do.cu.men.to)

sm

1. Qualquer produto de uma sociedade considerado como testemunho de uma época: A Muralha da China é um documento da inteligência e da cultura daquele povo.
2. Declaração escrita para servir de prova ou título (documento de identidade).; CERTIDÃO; CERTIFICADO

³⁷ Disponível em: <https://aulete.com.br>. Acesso em: 21 ago. 2021.

3. Qualquer texto de caráter público ou privado que preceitue, discuta, delibere ou solicite algo: Trata-se de um documento sobre o currículo do ensino secundário.
4. P.ext. Qualquer objeto que tenha valor documental (desenhos, escritos, fotografias, gravações, filmes etc.), que sirva para comprovar algum acontecimento, fato, algo que foi dito etc.
5. Inf. Arquivo com dados criado por programa, esp. os gerados por editores de textos [F.: Do lat. *documentum*, 'i'ensino' lição'.]

No contexto da Ciência da Informação, Buckland (1991) identifica três principais usos do termo "informação": *informação como processo*, *informação como conhecimento* e *informação como coisa*. **Informação como processo** refere-se ao ato comunicar um conhecimento ou notícia. É a ação de contar ou ser informado de algo. **Informação como conhecimento** refere-se ao que é percebido durante o processo de informar. É o conhecimento comunicado sobre algum fato, assunto ou evento. **Informação como coisa** refere-se à utilização do termo "informação" quando atribuído a objetos, quando estes transmitem conhecimentos e/ou informações.

Segundo Le Coadic (2004, p.5):

Documento é o termo genérico que designa os objetos portadores de informação. Um documento é todo artefato que representa ou expressa um objeto, uma ideia ou uma informação por meio de signos gráficos e icônico (palavras, imagens diagramas, mapas, figuras, símbolos), sonoros e visuais (gravados em suporte de papel ou eletrônicos).

Para Suzanne Briet (2006, p.10) define "documento" como “*qualquer signo físico ou simbólico, preservado ou registrado, com a intenção de representar, reconstruir ou demonstrar um fenômeno físico ou abstrato*”. Esta definição generaliza o conceito de documento a qualquer tipo de suporte, seja ele material ou digital.

2.2.2 Representação dos documentos

A recuperação de um documento depende da forma como foi representado. Na maioria dos sistemas de recuperação de informação essa representação é realizada pela atribuição de um conjunto de termos cujos significados remetem ao conteúdo informacional do documento.

Segundo Novellino (1996, p.38):

A principal característica do processo de representação da informação é a substituição de uma entidade linguística longa e complexa - o texto do documento - por sua descrição abreviada. O uso de tal sumarização não é apenas uma consequência de restrições práticas quanto ao volume de material a ser armazenado e recuperado. Essa

sumarização é desejável pois sua função é demonstrar a essência do documento. Ela funciona então como um artifício para enfatizar o que é essencial no documento considerando sua recuperação, sendo a solução ideal para organização e uso da informação

A representação de um determinado documento inclui os elementos descritivos que o identificam e o caracterizam em um acervo, assim como os elementos indicativos de seu conteúdo informativo, os assuntos por ele tratados. Tem como objetivo identificar cada documento de um determinado acervo, tornando-o "visível" para os usuários de um sistema de informação.

2.2.3 Usuário

Em um sistema de recuperação de informação, a tarefa do usuário é descrever e enunciar a sua necessidade de informação por meio de uma expressão de busca. A precisão dos resultados do sistema de recuperação de informação é dependente da qualidade da tradução da necessidade de informação do usuário em uma expressão de busca adequada. Segundo Riecken (2006, p.55):

O usuário e os conteúdos estão no centro das preocupações da CI [Ciência da Informação] e não as tecnologias. Entretanto, a CI pode valer-se de práticas de modelagem e notação oriundas da engenharia de software e da computação visando a facilitar a operacionalização em sistemas computacionais, suas bases de dados e redes, a partir dos modelos e representações conceituais.

2.2.4 Expressão de Busca

A expressão de busca (consulta) é o meio pelo qual o usuário comunica ao sistema a sua necessidade de informação. Geralmente é composta por um conjunto de termos ou palavras cujo significado representa a informação que necessita. Para enunciar expressões de busca que resultem documentos relevante, é necessário que o usuário tenha um mínimo de conhecimento do vocabulário do domínio ou assunto a qual se refere a sua necessidade de informação. A utilização de muitos termos vagos ou ambíguos pode gerar um excesso de documentos não relevantes.

2.2.5 Representação da Expressão de busca

Depois que o usuário define a sua expressão de busca, o sistema de recuperação de informação pode implementar diversos recursos que visem melhorar a expressão de busca inicialmente enunciada pelo usuário. Assim, uma nova expressão de busca pode ser construída

a fim de melhor representar a necessidade do usuário, resultando em conjunto de documentos mais relevantes.

2.2.6 Função de busca

No centro do processo de recuperação de informação está a função de busca, responsável pela comparação entre a representação da expressão de busca e a representação de cada um dos documentos. Na maioria dos sistemas de recuperação e dos mecanismos de busca Web esse processo resulta em uma espécie de lista com todos os documentos que contêm alguma relação com a necessidade do usuário.

2.2.7 Resultado da busca

No processo de recuperação de informação, o usuário utilizando termos do seu vocabulário pode encontrar uma lista de documentos que vai do mais relevante até o menos relevante, e a partir desse momento, ele pode analisar esse resultado, formular novas pesquisas alterando os termos utilizados na busca a fim de satisfazer a sua necessidade informacional de um ou mais documentos, por exemplo, 100% relevante em comparação com os termos usados.

2.2.8 Modelo de Recuperação de Informação

Baeza-Yates e Ribeiro-Neto (2011, p. 58) definem modelo de recuperação de informação como uma quadrupla:

$$[\mathbf{D}, \mathbf{Q}, \mathcal{F}, R(q_i, d_j)].$$

1. $\mathbf{D}$ é um conjunto composto por visões lógicas (representações) dos documentos no *corpus*;
2. $\mathbf{Q}$ é um conjunto composto de visões lógicas das necessidades de informação dos usuários;
3. $\mathcal{F}$ é um *framework* para a modelagem de representações dos documentos, consultas e seus relacionamentos;
4. $R(q_i, d_j)$ é uma função de ordenamento (*ranking*) que atribui um número real à relação entre uma representação da consulta q_i de $\mathbf{Q}$ e a representação de um documento d_j de $\mathbf{D}$.

Segundo Ferneda (2012, p. 20), um modelo de recuperação de informação é a especificação formal de três elementos: a *representação dos documentos*, a *representação da expressão de busca* e a *função de busca*.

A grande maioria dos modelos de recuperação de informação é baseada em disciplinas como a lógica, a estatística e a teoria dos conjuntos. Em um estudo sobre os modelos de recuperação de informação, Robertson (1977) justifica esse predomínio pelo fato de que a determinação de um modelo matemático geralmente pressupõe uma cuidadosa análise formal do problema e especificações de hipóteses, além de uma formulação explícita da forma como o modelo depende das hipóteses.

Este trabalho se limitará a apresentar resumidamente os três principais modelos de recuperação de informação, denominados modelos clássicos. São eles: modelo booleano, modelo vetorial e modelo probabilístico.

2.2.8.1 Modelo Booleano

No Modelo Booleano um documento é representado por um conjunto de termos que representa o seu conteúdo informacional. As buscas são formuladas por meio de uma expressão booleana composta por termos ou palavras ligados através dos operadores lógicos e apresentam como resultado o conjunto de documentos cuja representação satisfaz as restrições lógicas da expressão de busca. Em sua implementação básica o modelo booleano possui diversas limitações. Entre elas estão:

- Os termos que representam o conteúdo informacional de um documento possuem todos o mesmo peso, não sendo possível representar a importância relativa de cada termo na representação do documento. Da mesma forma, não é possível ao usuário definir a importância relativa de cada termo utilizado na expressão de busca.
- O resultado de uma busca booleana se caracteriza por uma simples partição do acervo documental em dois subconjuntos: os documentos que atendem à expressão de busca e aqueles que não atendem. Presume-se que todos os documentos recuperados possuem igual relevância, não possibilitando que os documentos possam ser ordenados.

Apesar de suas limitações, muitos sistemas se desenvolveram utilizando o modelo booleano como ponto de partida para a implementação de novos recursos de recuperação. Assim, pode-se dizer que o modelo booleano é o mais utilizado nos sistemas de recuperação de informação e nos mecanismos de busca da Web (FERNEDA, 2012, p.29).

2.2.8.2 Modelo Vetorial

No Modelo Espaço Vetorial, um documento é representado por um vetor numérico onde cada elemento representa o peso, ou relevância, do respectivo termo de indexação na representação do conteúdo informacional do documento. Da mesma forma que os documentos, uma expressão de busca também é representada por um vetor numérico onde cada elemento representa a importância (peso) do respectivo termo na representação da necessidade de informação do usuário (SALTON; WONG; YANG, 1975).

A utilização de uma mesma representação tanto para os documentos como para as expressões de busca permite calcular o grau de similaridade entre o vetor que representa uma determinada expressão de busca e cada um dos vetores que representam os documentos do acervo. Os valores da similaridade entre uma expressão de busca e cada um dos documentos do são utilizados no ordenamento dos documentos resultantes. Portanto, no modelo vetorial o resultado de uma busca é um conjunto de documentos ordenados pelo grau de similaridade entre cada documento e a expressão de busca.

2.2.8.3 Modelo Probabilístico

O Modelo Probabilístico foi proposto inicialmente por Maron e Kuhns (1960) e posteriormente explorado por diversos outros pesquisadores, tais como Robertson e Jones (1976). A ideia é tratar o processo de recuperação de informação como um processo probabilístico, já que ele é caracterizado por seu grau de incerteza. Assim, é mais realista pensar em uma probabilidade de relevância do que em uma pretensa relevância exata, como a utilizada nos modelos booleano e vetorial.

A necessidade de informação do usuário é representada por um ou mais termos, por meio de cálculos de probabilidade o sistema calcula, para cada documento, um valor numérico que representa a provável relevância do documento para a necessidade do usuário. Esse valor é utilizado para ordenar os resultados da busca. Tendo um primeiro conjunto de documentos, o usuário pode marcar alguns deles que considera relevantes. O conjunto de documentos marcados pode ser então submetido ao sistema, permitindo ao sistema recalcular a relevância de cada documento e fornecer resultados mais precisos. Esse processo, denominado *relevance feedback*, pode ser repetido até que o usuário se sinta satisfeito com os resultados.

Uma virtude do modelo probabilístico está em reconhecer que a atribuição de relevância é uma tarefa do usuário, pois é o único modelo que incorpora explicitamente o processo de *relevance feedback* como base para a sua operacionalização.

Os principais elementos envolvidos no processo de recuperação de informação (representação de documentos e expressões de busca) são geralmente de natureza linguística.

2.3 Processamento da Linguagem Natural

Assim, o PLN surge como uma possível solução aos problemas relacionados à recuperação de informação. PLN é um conjunto de técnicas computacionais para a análise de textos em um ou mais níveis linguísticos, com o propósito de simular o processamento humano da língua.

Segundo Faloutsos e Oard (1998), o desenvolvimento de sistemas de recuperação de informação que utilizam processos típicos do PLN possuem grande complexidade. Por esta razão, na maioria das vezes as técnicas de PLN são utilizadas apenas na melhoria do desempenho de algumas tarefas da recuperação de informação tradicional, como a indexação automática.

Atualmente existem sistemas de inteligência artificial e aprendizado de máquina que fornecem ferramentas e a capacidade computacional necessária para a construção de sistemas, não somente de PLN, mas também de Compreensão de Linguagem Natural (CLN) e de Geração de Linguagem Natural (GLN). Exemplos de usos hoje em dia são:

- **Sistemas de Tradução:** Sistemas que conseguem escutar um humano falando em alguma língua, fazer a tradução para outra língua e gerar uma resposta auditiva ou escrita, quase instantaneamente.
- **Assistentes de Voz:** Muitos sistemas foram desenvolvidos e a cada dia são mais utilizados por diversos níveis de usuários, como a Apple Siri, Google Assistente, Amazon Alexa, Samsung Bixby, Microsoft Cortana, entre tantos outros.
- **Sistemas de Correção Ortográfica:** Um sistema de correção contido em um editor de textos pode analisar e sugerir mudanças não apenas para palavras escritas erroneamente, mas pode corrigir erros gramaticais, de concordância, em frases, acentos, entre tantos outros; conseguindo quase entender a escrita humana e sugerindo a correção de frases inteiras que podem fazer mais sentido. Isso também

ocorre em sistemas de tradução, que podem sugerir frases melhores para o usuário em relação ao que este está tentando traduzir.

- *Chatbots*: Permite interagir com um usuário e por meio de um diálogo entender o que este necessita e sugerir resultados que podem ser desde respostas a perguntas, até mesmo a documentos e informações contidas em outras bases de dados ou documentos. No capítulo 3 serão abordados mais detalhadamente sobre o tema.

Nas subseções seguintes será discutida a utilização do PLN em alguns problemas clássicos da recuperação de informação. Segundo Lewis e Jones (1996), quase sem exceção, os métodos de PLN são utilizados em conjunto com os modelos clássicos.

2.3.1 Normalização de variações linguísticas

O reconhecimento de variações linguísticas das palavras de um texto proporciona maior precisão ao processo de indexação automática. A normalização linguística pode ser subdividida em três casos distintos: *morfológica*, *sintática* e *léxico-semântica* (Jacquemin; Klavans; Tzoukermann, 1997).

A **normalização morfológica** ocorre quando há redução dos itens lexicais através de *conflação* a uma forma que procura representar classes de conceitos. *Conflação* é a operação que combina a representação de dois ou mais termos em um único, reduzindo variantes de uma palavra a uma única forma.

Os procedimentos mais conhecidos para *conflação* são:

- *Stemming*: reduz uma palavra ao seu radical (*stem*) através da eliminação de afixos oriundos de derivação ou de flexão (Orengo; Huyck, 2001);
- *Redução à forma canônica*: processo também conhecido como *lematização*, que geralmente reduz os verbos ao infinitivo e os adjetivos e substantivos à forma masculina singular (Arampatzis, 2000).

A **normalização sintática** ocorre quando há a normalização de frases semanticamente equivalentes em uma forma única e representativa delas, "desempenhou com eficiência", "desempenho eficiente" e "eficiência em desempenho". Todas essas formas poderiam ser transformadas em "desempenho eficiente".

A **normalização léxico-semântica** ocorre quando são utilizados relacionamentos semânticos (como a sinonímia e hponímia) entre os itens lexicais para criar um agrupamento de similaridades semânticas, identificado por um item lexical que representa um conceito único: “estado emocional”, “estado afetivo” e “sentimento”. Esses termos poderiam ser reduzidos ao termo “sentimento”.

2.3.2 Identificação de termos compostos

Em um sistema de recuperação de informação geralmente termos compostos são identificados para que possam ser usados como termos de indexação, não se limitando à utilização de palavras ou termos isoladas.

Nos sistemas de recuperação de informação os termos compostos são geralmente identificados através de cálculos de coocorrência entre pares de palavras. A utilização da análise sintática permite identificar termos compostos mesmo quando as palavras que compõem o termo não são adjacentes ou não coocorrem com grande frequência.

Lewis e Jones (1996) notam que o grau de sofisticação do PLN poderia ser consideravelmente maior para as expressões de busca dos usuários do que para os documentos, pois a importância em entender quais são as necessidades do usuário, tendo em vista que geralmente os usuários de um sistema de recuperação de informação representam suas necessidades de informação por um número reduzido de termos ou palavras. Eventuais erros na interpretação das informações contidas em um documento podem ser compensados levando em conta outros termos extraídos do mesmo documento, o que não é possível para uma expressão de busca.

2.3.3 Resolução de ambiguidade

A ambiguidade é a propriedade que faz com que um objeto linguístico, seja uma palavra, um termo composto ou todo um texto, possa ser interpretado de modos diferentes. Quanto ao nível de processamento, existem dois tipos de ambiguidade: sintática e semântica. A ambiguidade sintática ocorre quando um item lexical pode pertencer a mais de uma classe gramatical, como “casa” que pode ser substantivo ou verbo. Outras causas da ambiguidade sintática são: mais de uma ligação possível do sintagma preposicional, como em “comprei um cofre com dinheiro”; a possibilidade de mais de uma coordenação ou conjunção, como em

“tenho amigos e parentes muito queridos”; ou a possibilidade de múltiplas combinações para substantivos compostos, como em “lareira da casa de pedras” (SMEATON, 1997).

As causas da ambiguidade podem ser dos seguintes tipos (BEARDON; LUMSDEN; HOLMES, 1991):

- **lexical**, que ocorre quando uma palavra pode possuir múltiplos significados;
- **estrutural**, quando é possível mais de uma estrutura sintática para a sentença, podendo ser: **local**, quando a ambiguidade pode ser resolvida dispensando o conhecimento do contexto em que ela ocorre; ou **global**, quando exige análise do contexto para sua resolução.

2.4 Recuperação de Informação como um Processo de Comunicação

A recuperação de informação permite resgatar documentos e informações que um usuário necessita e tem meios para garantir que a entrega será mais eficiente. Mas além desta finalidade, pode-se enquadrar o processo de recuperação de informação como um processo de comunicação entre o usuário e os documentos.

Para discutir esse fato, se precisa entender alguns conceitos para delimitar o campo de estudos e as áreas envolvidas. Alguns conceitos importantes ligados a essa temática permitem entender a ligação entre eles e as outras áreas. Aqui uma breve descrição de elementos relevantes (MEADOW *et al*, 2007, p. 37) apresentando uma visão para trazer algumas distinções necessárias, realçando que não podem ser atribuídas como as únicas definições corretas, mas na recuperação de informação, é preciso considerar alguns dos conceitos filosóficos ligados a essas palavras. Além de Dados e Informações, existem outros termos como Notícias, Conhecimento, Inteligência, Significado, Sabedoria, Relevância e Valor.

- **Dados:** Do latim *datum* (“aquilo que se dá”), um dado é considerado o valor de um atributo, na forma de uma sequência de símbolos, como dígitos ou letras. Um dado sozinho não traz significado, sentido ou conhecimento e não traz informação para o leitor.
- **Informações:** A informação é um dado que foi convertido em um contexto útil, ou seja, possui a conotação de dados avaliados, validados ou úteis. Pode-se dizer que dados que mudam a compreensão a respeito de alguma coisa sejam considerados informação, não importando se é um computador ou um ser humano. A informação

precisa ter um significado, se não produzir uma mudança no receptor não é uma informação. Como outras definições, pode até mesmo ter seu conceito vinculado à redução de incerteza, como por exemplo, “um tomador de decisão se depara com um conjunto de alternativas, não tem certeza de qual selecionar, e precisa de informações para fazer a escolha”.

- Notícias: Mesmo não sendo um termo importante para a compreensão da recuperação de informação, sua definição é próxima a das informações. A notícia é uma mensagem que se acredita ser verdadeira, mesmo inesperada. Se conhece o conteúdo de uma afirmação ou dado, não é uma novidade. Só pode ser considerada uma notícia se não souber o seu conteúdo e não o esperar. Um exemplo em um ditado “*O cão morde o homem não é novidade. Homem morde cachorro - isso é novidade.*”
- Conhecimento: Na maioria das definições o conhecimento tem a qualidade de ser um conjunto integrado de informações recebidas de fontes múltiplas, porém no uso geral, o conhecimento aparenta ser um grau mais alto de certeza ou validade do que uma informação, um ato de conhecer, ter ideia ou a noção de algo através de informações que lhe foram apresentadas. Enquanto uma informação é sugerida como um conjunto de fatos, o conhecimento requer compreensão. E até inteligível para as pessoas mais leigas equiparar dados e informações, mas comparar conhecimento com dados, para todos isso não parece adequado.
- Inteligência: Um dos conceitos é que a inteligência é a medida da capacidade de raciocínio, ou seja, a habilidade de manejar conhecimento e aplicá-lo na resolução de problemas. A maior relevância deve ser dada ao uso, porque para propósitos práticos, inteligência é informação.
- Significado: Cita o fato de que este descritor relacionado à informação é o mais difícil de ser definido. Sugerir que palavras são símbolos de coisas, ações, qualidades, relacionamentos etc., acaba sendo considerada uma simplificação muito grosseria para tal. “O significado completo de uma palavra não aparece até que ela seja colocada em seu contexto, e o contexto pode servir a uma função extremamente sutil - como nos trocadilhos ou duplo sentido. E mesmo então, o “significado” dependerá do ouvinte, do falante, de toda a sua experiência da língua, do conhecimento mútuo e de toda a situação. (MEADOW *et al*, 2007, p. 42, tradução

nossa)”³⁸. Para a recuperação de informação, seria muito mais pacífico se cada palavra tivesse seu próprio significado e fosse imutável, porém na realidade isso não acontece, nem mesmo se pessoas lerem o mesmo texto, ainda poderão interpretá-lo de formas diferentes. Sobre um mesmo assunto, usuários pensam de forma diferente no momento de expressar o significado das informações que precisam recuperar.

- Sabedoria: Pode ser descrita como um dom de uma pessoa que não tem mais dados ou informações do que as outras, mas cujas declarações têm uma maior probabilidade de serem aceitas como verdadeiras por grupos de usuários. Ele consegue reconhecer relacionamentos entre observações, que até aquele momento, não haviam sido observadas como relacionadas e fornecer informações a respeito de questões importantes.
- Relevância e Valor: Estes dois termos fazem parte e são de suma importância para grande parte dos sistemas de recuperação de informação. O mais importante para um sistema desse tipo é se uma informação que foi recuperada é útil ou não para o usuário que a está procurando, muito mais importante do que se o que foi recuperado é verdadeiro ou falso, pois isso, de certa forma, depende de o usuário determinar. Um pouco adiante serão apresentadas mais informações sobre estes termos e os relacionados.

Quando um usuário precisa de uma informação sobre qualquer assunto, ele recorre a um sistema que pode auxiliá-lo nesse problema, visto que ele tem informações sobre o que precisa e um desejo do que encontrar. Para tal, formula uma expressão de busca e a submete para ser utilizada pelo sistema para fornecer os termos necessários para uma pesquisa inicial e em seguida apresenta ao usuário os resultados. Para o operador que busca uma informação, ele não está interessado como o resultado será mostrado, não importa se são documentos, *links*, textos etc., ele quer que a resposta atenda a sua pergunta e satisfaça sua necessidade, se isso não ocorrer, ele pode reformular sua questão de busca e submeter novamente, recomeçando esse “diálogo” com o sistema, quantas vezes forem necessários. Ou seja, “parte-se da premissa de que o usuário está interessado em recuperar informação sobre um determinado assunto e não documentos, embora a informação esteja alocada nos documentos da qual a informação está

³⁸ The full meaning of a word does not appear until it is placed in its context, and the context may serve an extremely subtle function—as with puns, or double entendre. And even then the “meaning” will depend upon the listener, upon the speaker, upon their entire experience of the language, upon their knowledge of one another, and upon the whole situation.

registrada” (FERNEDA; DIAS, 2013). Essa interação que diferencia um sistema de recuperação de informação de um gerenciador de banco de dados.

Ao tratar desses conceitos na busca de recuperar informações relevantes para um usuário, houve avanços expressivos nas tecnologias de comunicação e de informação, acarretando uma aproximação das disciplinas de comunicação e informação a quase todas as áreas do conhecimento, que antes pareciam tão distantes, mas que na prática se mostraram tão próximas. A insistência em divergir leva a uma confusão que pode levar ao não aumento do conhecimento por parte dos pesquisadores dessas áreas, não importando se das áreas da medicina, enfermagem, estatística, negócios, artes etc. Ao tratar os conceitos de comunicação e informação como meios, e não como fim, essas tecnologias se tornaram vitais para as áreas da Ciência da Comunicação e da Ciência da Informação e puderam contribuir de forma muito positiva nas pesquisas de outras áreas e fornecer meios e conhecimento para que os pesquisadores pudessem avançar a tecnologia na busca da melhoria da vida das pessoas (CIBANGU, 2015).

Veja-se o conceito de comunicação, visto que é importante para se entender a relação entre o usuário e a informação cujo intermédio se dá pelos sistemas de recuperação de informação. Segundo o Dicio, Dicionário Online de Português³⁹ o conceito de comunicação é:

Significado de Comunicação

substantivo feminino

Ação ou efeito de comunicar, de transmitir ou de receber ideias, conhecimento, mensagens etc., buscando compartilhar informações.

...

Etimologia (origem da palavra comunicação). Do latim communicatio, "ação de participar".

Já de acordo com o dicionário da língua portuguesa na Internet, Aulete Digital⁴⁰, o conceito é:

(co.mu.ni.ca.ção)

sf.

Conceito, capacidade, processo e técnicas de transmitir e receber ideias, mensagens, com vistas à troca de informações, instruções etc.: A comunicação é um pré-requisito para a formação e consolidação de uma sociedade.

...

[Pl.: -ções.]

³⁹ Disponível em: <https://www.dicio.com.br/comunicacao>. Acesso em: 21 ago. 2021.

⁴⁰ Disponível em: <https://aulete.com.br/comunicação>. Acesso em: 21 ago. 2021.

[F.: Do lat. communicatio, onis.]

Outro que possui uma definição para a palavra comunicação é o Priberam Dicionário⁴¹, que é:

co·mu·ni·ca·ção
(latim communicatio, -onis)
substantivo feminino
1. Informação; participação; aviso.
2. Transmissão.
3. Notícia.
4. Passagem.
5. Ligação.
6. Convivência.
7. Relações.
8. Comunhão (de bens).

De acordo com estas definições, pode-se concluir que o termo comunicação se refere ao ato de se comunicar, um processo de transmissão e recepção de ideias ou mensagens, entre um receptor e um transmissor com o intuito de trocar informações, entre outras afirmações. No contexto da recuperação de informação, Ferneda e Dias (2015, p. 1) esclarecem que “um sistema de recuperação de informação é um ambiente linguístico mediador da comunicação entre um estoque de informação e seus requisitantes. Sua eficiência depende de um controle adequado da representação dos documentos e das requisições de seus usuários.”

Vieira (1994) resume essa relação dialógica entre o usuário e um sistema de recuperação de informação,

A recuperação de informações é um processo de comunicação em que emissor e receptor interagem para atender a uma necessidade de informação. Ao fazer a pergunta ao sistema, o homem funciona como um emissor e o computador como um receptor. Por outro lado, o computador, ao apresentar sua resposta, passa a ser o emissor e o homem o receptor. Essa interação só é viável por meio do uso da linguagem. (VIEIRA, 1994, p. 6, tradução nossa) ⁴²

⁴¹ Disponível em: <https://dicionario.priberam.org/comunicação>. Acesso em: 21 ago. 2021.

⁴² La recuperación de información es un proceso de comunicación donde se relaciona el emisor y receptor con el fin de cubrir una necesidad de información. Al hacer la pregunta al sistema el hombre funciona como emisor y el ordenador como receptor. En contrapartida el ordenador al presentar su respuesta pasa a ser el emisor y el hombre el receptor. Esta interacción solo es viable a través del uso del lenguaje.

Ao prover para o sistema de recuperação de informação uma quantidade maior de fontes de dados e permitir um diálogo (comunicação) entre o usuário e o sistema, Pöttker (2017) percebeu um aumento na eficiência do sistema.

Portanto, comprova-se a tese de que um sistema de recuperação de informação é mais eficiente quando amplia seu escopo de recuperação a partir de diferentes fontes de dados (ROA) e permite a inferência do usuário no julgamento da informação que lhe foi retornada. Assim, acredita-se que um sistema de recuperação de informação retorna resultados mais relevantes quando possibilita estabelecer um processo dialógico entre o usuário e o sistema de recuperação de informação (PÖTTKER, 2017, p. 176)

Como apresentado neste capítulo, a recuperação de informação é um processo eminentemente linguístico e dialógico. Um sistema de recuperação de informação é um elemento mediador desse diálogo no qual um usuário enuncia a sua necessidade de informação e obtém como resposta um conjunto de documentos nos quais possivelmente ele encontrará a informação que necessita. Como em todo processo comunicativo, essa interação entre usuário e sistema está propensa a ruídos, comum à toda linguagem: subjetividade, ambiguidade e imprecisão. Para a redução desses ruídos inerentes à comunicação, esse trabalho propõe a utilização dos conceitos e tecnologias de construção de *chatbots* para a solução de parte desses problemas. Propõe-se assim uma comunicação bilateral, na qual se pode notar uma reciprocidade entre o emissor e o receptor, mesmo que um desses elementos seja um *software*.

3.

CHATBOTS

Neste capítulo serão descritas algumas tecnologias que vêm sendo utilizadas por diversas áreas como interfaces que simulam agentes conversacionais com a capacidade de interagir com os usuários: como funcionam e como podem ser construídos.

3.1 Definição de *Chatbot*

Os sistemas de diálogo baseados em texto ou voz, também chamados de agentes de conversação, *chatbots* ou *chatterbots*, estão se tornando populares, não só por parte das grandes empresas, mas também pelos usuários de Internet e de dispositivos móveis. E uma das razões dessa popularidade é a sua capacidade de interagir de forma inteligente, que vem melhorando muito nos últimos tempos devido aos avanços significativos nas tecnologias de *hardware* e inteligência artificial (BARTL; SPANAKIS, 2017).

O termo *chatbot* é formado pela junção das palavras da língua inglesa: *chat* (bate-papo) e *bot*, abreviação da palavra *robot* (robô). É um programa de computador que possui a capacidade de imitar uma conversa inteligente entre um humano e um computador, permitindo ainda, examinar e até mesmo influenciar o comportamento do usuário desse sistema. Por meio da construção de perguntas e respostas, ele funciona como um mediador/interface na comunicação de um sistema computacional com os seus usuários na busca de sanar as necessidades de informações destes. O sistema tem a capacidade de formular perguntas de maneira inteligente para o usuário por meio de linguagem natural, forma que os humanos se comunicam entre si, além disso, essa comunicação pode ser de forma textual ou falada. Conforme a interação vai progredindo, o sistema é capaz de processar todos os termos do diálogo e construindo respostas, e novas perguntas, de maneira que tenha a capacidade de

entender e ser entendido pelo usuário, podendo repetir este processo quantas vezes for necessário (ABDUL-KADER; WOODS, 2015).

O conceito de *chatbot* evoluiu muito desde que Alan Turing publicou seu artigo em 1950. Na sua publicação "*Computing Machinery and Intelligence*", questionou se as máquinas poderiam pensar. Propôs um tipo de avaliação para identificar se um computador poderia ser confundido com um ser humano e dessa forma ser considerado inteligente, ele deu o título "*The Imitation Game*", em seu artigo, e que hoje é conhecido como "Teste de Turing" (TURING, 1950).

É possível classificar estes assistentes robóticos em vários tipos, de acordo com Saceano (2000), podem ser: *searchbots*, que podem auxiliar a procura na Internet, *mailbots*, sistemas que podem interagir utilizando *e-mails*, *modbots*, moderadores de fóruns na Internet e os *chatbots* que podem simular diálogos utilizando textos ou sistemas multimídia.

Classificando os *chatbots* em 3 diferentes gerações, Rothermel e De Souza Domingues (2007) descrevem a primeira geração como baseada em relacionamento entre padrões e regras gramaticais, por não registrar as conversas, não têm acesso ao histórico dos diálogos anteriores, e como exemplo, o robô ELIZA. Na segunda geração, como exemplo o robô JULIA, a inteligência artificial e as redes neurais foram incorporadas nas suas regras de funcionamento. Na terceira geração a utilização de linguagens de marcação como o AIML (*Artificial Intelligence Markup Language*) para a criação de uma base de conhecimento, como exemplo, o ALICE.

Atualmente a utilização de assistentes pessoais do tipo *chatbot*, entre outros, está ficando cada dia mais presente em vários domínios da sociedade. Estes sistemas cognitivos trabalham como especialistas que podem fornecer respostas para a resolução de quase todos os tipos de problemas e tarefas rotineiras. Além dos meios tradicionais de contato com os usuários, os *bots* podem fornecer canais adicionais para as perguntas frequentes, linhas diretas e até mesmo fóruns, que de forma interativa, podem atender aos usuários (LOMMATZSCH; KATINS, 2019).

Os *chatbots* precisam suportar diálogos mais longos, na forma de uma pesquisa interativa, e orientar os usuários na busca de soluções para problemas mais complexos. Além disso, ao lidar com uma pesquisa em sistemas de informações deve ser capaz de processar consultas utilizando uma variedade de palavras-chave e frases complexas utilizando linguagem natural (LOMMATZSCH; KATINS, 2019).

A respeito da cronologia e história dos *chatbots*, Sumit Raj (2019) em seu livro descreve o quanto se avança desde o início, enumerando a maioria dos exemplos mais conhecidos:

1950 - O teste de Turing foi desenvolvido por Alan Turing. Ele testava a habilidade de uma máquina exibir comportamento inteligente equivalente ao, ou indistinguível do, de um humano.

1966 - Eliza, o primeiro chatbot, foi criado por Joseph Weizenbaum, projetada para ser uma terapeuta. Ela simulava uma conversa usando uma metodologia “comparação de padrões” e substituição que dava aos usuários a impressão de compreensão por parte do *bot*.

1972 – Parry, um programa de computador do psiquiatra e cientista de Stanford Kenneth Colby, modelava o comportamento de esquizofrenia paranoide.

1981 – O *chatbot* Jabberwacky foi criado pelo programador britânico Rollo Carpenter. Ele começou a ser criado em 1981 e foi lançado na Internet em 1997. O objetivo desse *chatbot* era “simular uma conversa humana natural de maneira interessante, agradável e bem-humorada”.

1985 – O robô de brinquedo *wireless*, Tomy Chatbot, repete mensagens gravadas em sua fita.

1992 – Dr. Sbaitso, um *chatbot* criado pela Creative Labs para MS-DOS, “conversava” com o usuário como se fosse um psicólogo em voz digitalizada. Palavrões repetidos e entradas mal elaboradas fornecidos pelos usuários fizeram o Dr. Sbaitso “travar” em um “ERRO DE PARIDADE” antes que pudesse se reinicializar.

1995 – A.L.I.C.E. (*Artificial Linguistic Internet Computer Entity*) foi desenvolvida pelo ganhador do Prêmio Nobel Richard Wallace.

1996 – Hex, desenvolvido por Jason Hutchens, foi baseado na Eliza e ganhou o Prêmio Loebner em 1996.

2001 – Smarterchild, um *bot* inteligente desenvolvido pela ActiveBuddy, foi amplamente distribuído em trocas de mensagens instantâneas globais e redes SMS. A implementação original evoluiu rapidamente para dar acesso instantâneo a notícias, previsão do tempo, informações do mercado de ações, horários de sessões de cinema, listagem de páginas amarelas e dados detalhados de esportes, além de fornecer várias ferramentas (assistente pessoal, calculadoras, tradutor etc.).

2006 - A ideia do Watson foi concebida em um jantar, ele estava sendo projetado para competir no programa de TV “Jeopardy”. Em sua primeira tentativa, só acertou cerca de 15% das perguntas, mas depois conseguiu vencer concorrentes humanos regularmente.

2010 – O Siri, um assistente pessoal inteligente, foi lançado como um aplicativo do iPhone e depois integrado como parte do iOS. Ele é resultado de um esforço do Centro de Inteligência Artificial da SRI *International*. Seu mecanismo de reconhecimento de fala foi fornecido pela Nuance Communications; o Siri usa tecnologias avançadas de *machine learning* para funcionar.

2012 – O Google lançou o *chatbot* Google Now. Originalmente seu codinome era “Majel” em homenagem a Majel Barret, esposa de Gene Roddenberry e voz dos sistemas de computador da franquia Star Trek; também recebeu o codinome de “Assistant”.

2014 – A Amazon lançou a Alexa. A palavra “Alexa” tem uma consoante forte no X, logo, pode ser reconhecida com maior precisão. Essa foi a principal razão para a Amazon escolher esse nome.

2015 - Cortana, assistente virtual criada pela Microsoft. A Cortana pode criar lembretes, reconhecer voz natural e responder a perguntas usando informações do mecanismo de busca Bing. Seu nome vem de uma personagem de inteligência artificial fictício da série de videogames Halo.

2016 - Em abril de 2016, o Facebook anunciou uma plataforma de *bot* para o Messenger que incluía APIs de construção de *chatbots* para a interação com os usuários. Melhorias posteriores incluíram *bots* que participam de grupos, telas de pré-visualização e recursos de leitura de QR por intermédio da funcionalidade de câmera do Messenger para levar os usuários diretamente para o *bot*.

Em maio de 2016, o Google revelou seu *bot* ativado por voz concorrente da Amazon Echo chamado Google Home na conferência de desenvolvedores da empresa. Ele permite aos usuários emitir comandos de voz para interagir com vários serviços.

2017 – O Woebot um agente conversacional que nos ajuda a monitorar nosso humor, a nos conhecer melhor e a nos sentir bem. Ele usa uma combinação de técnicas de NLP, habilidades psicológicas (terapia cognitivo-comportamental [TCC]), ótima redação e senso de humor para tratar a depressão (RAJ, 2019, p. 30).

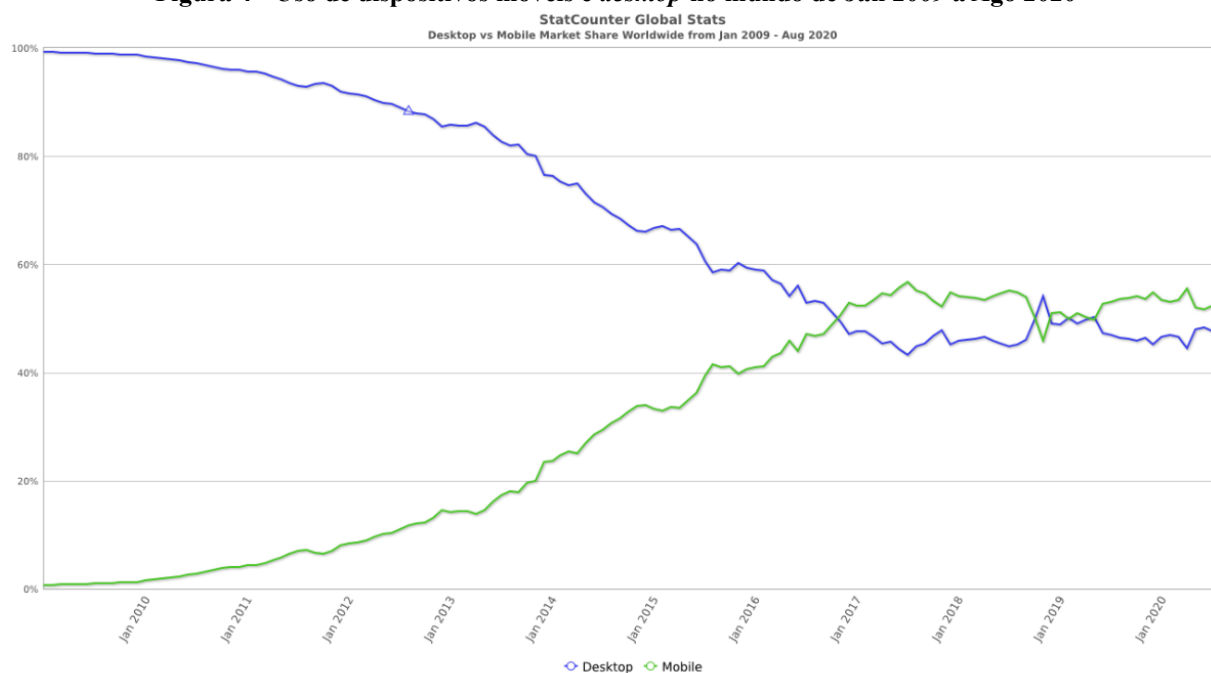
Atualmente, os *chatbots* mais modernos são desenvolvidos por grandes empresas que investem muito dinheiro em pesquisas e aperfeiçoamento, empresas como Amazon, Apple, Microsoft, Facebook, entre muitas, utilizam dos últimos avanços em sistemas de aprendizagem de máquina para aplicá-los em sistemas de recuperação de informações que utilizando de pesquisas pela Internet podem responder a muitas perguntas. Outros modelos adotados são os que utilizam técnicas de tradução automática de estatísticas para tentar traduzir as frases de entrada em respostas de saída ou mesmo a utilização de redes neurais que codificam e decodificam entradas em respostas (CAHN, 2017).

Outro fato que vem contribuindo para o aumento da utilização de sistemas conversacionais é a quantidade de pessoas que utilizam telefones inteligentes e até mesmo a diminuição das que utilizam computadores *desktop* ou *laptop*, grande parte dos sistemas funcionam até melhor para auxiliar os usuários quando estes o utilizam através de um telefone, visto que muitas pessoas não conseguem ficar longe de seus aparelhos e outros até mesmo nem possuem computadores em suas casas, de forma que todos os aspectos da vida cotidiana das pessoas estão sendo gerenciadas por esse equipamento, o que faz muito sentido investir em tecnologias que possam auxiliar o usuário para encontrar as suas informações.

Na Figura 4 tem-se um gráfico que mostra a porcentagem de participação no mercado de acesso à Internet, onde a utilização de computadores vem caindo com o passar dos anos e ao mesmo tempo um aumento de aparelhos móveis na utilização de recursos informacionais no mundo. Além disso, no final de 2016, a quantidade de aparelhos móveis já supera a quantidade de computadores neste tipo de utilização. Em dezembro de 2021, o número mais atualizado da participação no mundo é de 54,86% para dispositivos móveis, 42,65% para computadores e 2,49% de *tablets*, números disponibilizados diariamente por Statcounter (2022), que descreve a sua metodologia de pesquisa para as estatísticas, como baseadas em dados agregados e

coletados pelo Statcounter em uma amostra superior a 10 bilhões de visualizações de página por mês coletadas em toda a rede Statcounter de mais de 2 milhões de *sites*. As estatísticas são atualizadas e disponibilizadas todos os dias, no entanto, estão sujeitas a testes e revisão de garantia de qualidade por 45 dias a partir da publicação.

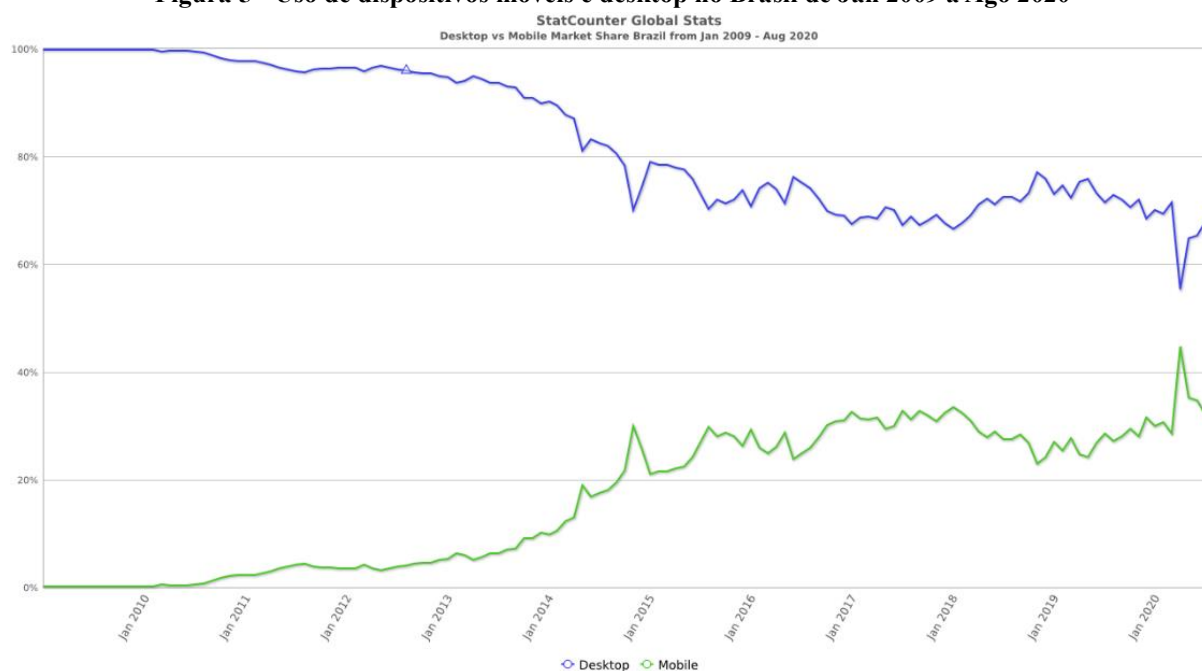
Figura 4 - Uso de dispositivos móveis e *desktop* no mundo de Jan 2009 a Ago 2020



Disponível em: <http://gs.statcounter.com>. Acesso em: 02 jan. 2022.

Já a Figura 5, mostra que no mesmo período no Brasil, a quantidade destes equipamentos segue a mesma tendência que no resto do mundo, com uma diminuição a cada ano da utilização de computadores para acesso à Internet e sendo substituído por dispositivos móveis. Em dezembro de 2021 o número de utilização por meio de computadores foi de 50,66%, enquanto os dispositivos móveis chegaram a 48,6% e de *tablets* de 0,73% (STATCOUNTER, 2022).

Figura 5 - Uso de dispositivos móveis e desktop no Brasil de Jan 2009 a Ago 2020



Disponível em: <http://gs.statcounter.com>. Acesso em: 02 jan. 2022.

A partir de um cenário assim descrito, com as tecnologias de acesso à disposição dos usuários e o interesse das empresas em construir e disponibilizar novas interfaces de acesso, veja-se as principais características dos *chatbots*.

3.2 Características dos *chatbots*

Ao questionar quais aptidões um sistema cognitivo na forma de um agente conversacional precisa possuir, Liu *et al* (2017) descreve seis habilidades necessárias para que o sistema desempenhe essa função de uma maneira mais eficiente:

1. Habilidade de filtrar informações supérfluas, como por exemplo, preenchimentos, pausas, falsos inícios etc.
2. O sistema deve produzir uma resposta adequada e determinar a necessidade de uma tomada de decisão, por exemplo, se deve mostrar resultados, escolher dentre uma hierarquia ou pedir uma consulta com acompanhamento etc.
3. Decidir se deve apresentar uma lista de resultados classificados ou a reunião das respostas para apresentar apenas uma resposta mais completa.
4. Que mantenha um histórico das conversas bem-sucedidas, sustentando os objetivos das buscas, além do estado atual do entendimento do sistema.

5. A fim de processar corretamente as consultas dos usuários, o sistema deve possuir um conhecimento geral sobre o mundo externo para explorar informações contextuais durante o processo.
6. Deve possuir uma personalidade e uma responsabilidade moral para interagir nos casos de consultas de teor emocional ou sensíveis aos usuários.

Quando um *bot* e um usuário iniciam uma conversa, algumas possíveis atitudes podem se desenrolar, nesse processo é necessário extrair algumas informações específicas para que o sistema possa responder ou interagir com esse usuário. Azzopardi *et al.* (2018) descrevem algumas premissas para a utilização do sistema de forma positiva e produtiva, como:

- A intenção de ajudar o usuário de forma bastante cooperativa, buscando servir aos interesses e desejos do usuário e não apenas na busca de informações para seu processamento e construção de aprendizado.
- Que durante a interação não aconteça diálogos ociosos para a busca das informações, minimizando assim o esforço nessa conversa, evitando sobrecarregar a capacidade cognitiva do usuário fazendo perguntas excessivas ou oferecendo resultados ou opções irrelevantes.
- Procurar oferecer o máximo possível de opções relevantes para o usuário no tempo disponível para fazer a conversa, compreendendo o contexto da pesquisa, orientando a exploração do tema e oferecendo escolhas de acordo com os desejos do usuário e sua capacidade cognitiva de formular as perguntas.

Ainda a respeito das funções esperadas que sejam atribuídas aos *chatbots*, Cahn (2017) os descreve como sistemas *online* de diálogo humano-computador com linguagem natural e que as funções esperadas de *chatbots* modernos seriam como agentes dialógicos, racionais e encarnados. Como agente dialógico deve possuir a função de compreensão, ou seja, deve entender o usuário, pois ao utilizar ferramentas de processamento de linguagem natural pode entender as informações fornecidas pelo usuário e gerar respostas apropriadas para mais variadas situações. Ao tratar o *chatbot* como um agente racional, é suposto que este armazene informações específicas do contexto do diálogo, pois ao ter acesso à uma base externa de conhecimento e de senso comum pode fornecer informações às perguntas do usuário de maneira mais acertada. Para que a utilização desses agentes fosse mais agradável a presença do ser humano, a função de estar encarnado se dá quando se pode tratá-los com uma personalidade que poderia até ser atribuída a outro ser humano, um exemplo disso, é que desde os primeiros sistemas de conversação se davam nomes de pessoas como Eliza, Alice, Charlie, entre outros,

para que dessa forma pudesse construir um grau de confiança com os usuários e dar a impressão de humanidade ao sistema.

Na perspectiva de avaliar o desempenho de um *chatbot*, Cahn (2017) comenta sobre alguns processos, como por exemplo, para a recuperação da informação eles têm funções específicas, existindo como assistentes virtuais, sistemas de perguntas e respostas e *bots* específicos de domínio. Para medir a precisão e a eficácia, um avaliador deve fazer perguntas e solicitar informações ao *chatbot*, dessa forma é possível avaliar a precisão das respostas.

Já do ponto de vista da experiência de um usuário, o objetivo do *bot* é utilizar todos os meios para maximizar a satisfação do requerente. Quando os pesquisadores analisam os usuários, podem se utilizar de meios como questionários que poderão ser utilizados para classificar os *bots* com base nas suas habilidades e satisfação no uso. A respeito da sua capacidade de comunicação, do ponto de vista linguístico, os *bots* precisam ser avaliados sobre a sua capacidade de gerar frases completas e corretas gramaticalmente e cognitivamente. Além disso, é necessária avaliação da sua inteligência para que ele possa convencer um ser humano que ele é confiável e eficaz no trabalho de recuperar as informações.

3.3 Tipos de *Chatbots*

Tem se verificado nos últimos tempos um aumento na utilização desses sistemas de diálogo e agentes conversacionais, sendo utilizados também por grandes, médias e pequenas empresas para estreitar as relações com seus clientes a fim de reduzir o tempo de resposta às perguntas destes e para reduzir os custos desses sistemas de contato. Podem ser encontrados em empresas do setor privado e no setor público, tanto na utilização de assistentes virtuais alimentados por voz, que se verá mais adiante, e no atendimento de agências de telecomunicações, quanto no provisionamento de transações nos bancos, reservas de hotéis e restaurantes no turismo, provisão de notícias na mídia, comércio, mercado de ações e seguros. Outro uso que tem ganhado muitos utilizadores é a utilização de *chatbots* em plataformas de redes sociais para interação com o público.

No processo de desenvolvimento desses sistemas, algumas abordagens podem ser usadas, como descrito por Makhalova, Ilvovsky e Galitsky (2019), podem ser classificados como *chatbots* sociais ou orientados a tarefas. *Chatbots* sociais podem construir conversas com os humanos das mais diferentes formas, simulando a espontaneidade. Para conseguir tal façanha, ele precisa obter um grande conjunto de históricos de diálogos de treinamento e com

essas informações, construir uma rede de aprendizagem para “aprender como conversar”. Quando se trata de um conhecimento específico de um domínio, um *chatbot* orientado a tarefas pode contribuir, pois utilizando de processamento de linguagem natural e aprendizagem de máquina pode manter uma conversa de forma natural a cerca de um domínio, que pode ser como a recuperação de informações de restaurantes, reservas de voos ou suporte automático a clientes.

Conforme Shevat (2017), em seu livro, descreve a diferença entre os tipos de *bots* e que devido a essa grande variedade e a possibilidade de compreendê-los é possível adquirir a capacidade de escolher o tipo certo para cada caso e explorar as melhores maneiras de aplicá-los ao seu serviço, marca ou produto. Eles podem ser classificados como *bots* pessoais ou de equipe, *superbots* ou *bots* de domínios específicos, *bots* de negócios ou *bots* de consumidores, *bots* de voz ou texto, novos *bots* de rede ou *bots* de sistemas legados.

Prossegue descrevendo estes tipos e suas principais características:

Um *bot* pessoal funciona como um assistente pessoal. Ao comunicar-se com o usuário diretamente ele possui foco porque o seu atendimento é unitário, ou seja, um usuário de cada vez. Já o *bot* de equipe existe para facilitar a comunicação em um trabalho que envolve muitas pessoas, ou seja, ele pode falar com vários usuários ao mesmo tempo, em uma configuração como se fosse um grupo para que esses vários usuários possam trocar informações importantes entre si. Um *bot* pessoal trabalha com um usuário e um contexto de cada vez, enquanto o de equipe pode manipular uma quantidade muito maior de contextos para vários usuários.

Para representar um único produto/serviço/marca é recomendada a utilização de um *bot* de domínio específico, como por exemplo, um *bot* de viagens aéreas, pois ele pode ajudar em tudo o que tem a ver a este domínio, desde a reserva de voos, o fornecimento de alertas para viagens e todo o controle até a chegada. Um outro tipo é o *superbot* que pode ajudar o usuário a utilizar diversos serviços por meio dele. Um mesmo sistema se conectando à uma infinidade de outros sistemas de forma a estender as suas funcionalidades, um bom exemplo é o Assistente do Google, a partir da qual é possível criar um horário em uma agenda, buscar dados na Internet, controlar uma lâmpada, procurar músicas ou abrir outros programas, como de mapas. A vantagem de utilizar um sistema assim é que o usuário não precisa lidar com vários sistemas para conseguir acessar as informações que deseja. Já a vantagem de se utilizar um *bot* de domínio específico é que ele tem conhecimento de toda arquitetura do domínio e pode facilitar

o acesso aos serviços para usuário, eliminando tempo perdido na busca através de menus para encontrar o que precisa.

Os *bots* de negócios e os *bots* de consumidores são diferentes em muitos aspectos, pois possuem finalidades diferentes e se envolvem com os usuários de maneiras diferentes. Os *bots* de negócios podem atuar como um gerente que coordena os membros de uma equipe no fluxo de trabalho, fazendo que a comunicação seja direta e com foco em fazer as coisas de maneira mais produtiva. Ele pode interagir com vários funcionários de modo a garantir que os processos sejam criados, aprovados e enviados dentro dos prazos. Já os *bots* de consumidores são mais orientados para a experiência do usuário, onde seus objetivos estão em entreter e facilitar o comércio, também podem manter contato com as marcas favoritas dos usuários, os atualizando sobre as novidades, de olho no bem-estar e não no fluxo de tarefas.

A respeito da maneira que os usuários interagem com estes sistemas, pode-se dizer que podem ser através de voz ou texto. Os *bots* que trabalham por texto, normalmente estão presentes em aplicativos de bate-papo, já os que utilizam a voz como principal sistema de conversação, atualmente, são utilizados em grande maioria nos assistentes instalados em equipamentos e telefones celulares. Para utilização por voz normalmente é necessário que o usuário aperte um botão ou utilize um comando de voz ajustado anteriormente entre o aplicativo e o usuário.

Quando uma empresa já utiliza um sistema de gerenciamento de suas atividades ela pode ter interesse em criar um *bot* de integração nos sistemas legados. Pode começar com pequenas funcionalidades em que esse aplicativo pode extrair do sistema da empresa e fornecer na forma de uma conversa com os usuários, podendo com o tempo, fornecer mais funcionalidades atribuídas a ele, tornando-o mais uma interface para acesso ao sistema, devido ao fato que esse sistema principal não deve deixar de existir tão cedo. Extraindo informações deste sistema ele pode construir seu contexto e utilizá-los na conversa, utilizando apenas dados importantes e relevantes, dentre a grande quantidade de informações que podem estar disponíveis no produto. Um novo tipo de *bot* chamado pelo autor de novos *bots* de rede pretendem ser uma nova proposta de desenvolvimento de interfaces, onde a conversação é a maneira de se comunicar, desde o início, com o sistema de controle do negócio. O principal desafio dessa modalidade é o desenvolvimento do *design* de forma a permitir o acesso a todas as funcionalidades.

Ainda de acordo com Shevat (2017), “*bots* são uma nova interface de usuário que pode ser implementada de várias maneiras e resolver diferentes tipos de negócios e casos de uso do consumidor. Eles podem estender e aumentar os serviços existentes e expor novos.”

3.4 Diferença entre Assistente Digital e *Chatbot*

Ao comparar esses dois programas muitas pessoas podem até generalizar e atribuir as mesmas funções a ambos, mas existem diferenças entre eles, como explicado por Oracle, a definição mais simples de *chatbot* é que são “programas de computador que simulam e processam conversas humanas (escritas ou faladas), permitindo que as pessoas interajam com dispositivos digitais como se estivessem se comunicando com outra pessoa” (ORACLE, 2020). E como exemplo, comprar um produto numa loja usando um *chatbot* ou requerendo um serviço de corrida ou carona, onde a conversa é bastante simples e direta.

Já a utilização de um assistente digital requer um pouco mais de tecnologia para alcançar seus objetivos. Ao trabalhar com várias fontes de dados e com o foco de colocá-las em contexto e trabalhando juntas, dependendo da finalidade, requer habilidades diferentes. Para processar o que o usuário está dizendo, ou escrevendo, é preciso capacidade avançada de processamento de linguagem natural, pois neste caso, o usuário pode utilizar frases complexas e é necessário separar o que ele disse para o completo entendimento e a geração de respostas mais precisas, pois uma mesma frase pode conter mais de uma pergunta ou afirmação. Além do processamento da linguagem, esses assistentes usam a inteligência artificial e o aprendizado de máquina para estudar as preferências com base nas interações anteriores, e a partir disso, pode prever seu comportamento para oferecer recomendações baseadas neste histórico. O resultado é uma experiência mais personalizada e condizente com as necessidades dos usuários (ORACLE, 2020).

Ainda complementa, resumindo as diferenças entre assistentes digitais e *chatbots*, como sendo o assistente digital um tipo de *chatbot* avançado que pode trabalhar com diálogos mais complexos e de maneira conversacional mais elaborada. “Embora todos os assistentes digitais sejam *chatbots*, nem todos os *chatbots* são assistentes digitais” (ORACLE, 2020).

Quando uma empresa investe em assistentes digitais procura a melhoria da eficiência e assistência aos envolvidos, além disso pode economizar quando conecta a empresa aos clientes de forma mais eficiente e rápida liberando a equipe para desempenhar trabalhos mais críticos, fornecer um melhor serviço conversando com os clientes por meio de aplicativos de mensagens

que já estão acostumados, evitar erros quando pode antecipar eventos futuros e informar aos usuários sobre prazos e inconsistências, escalar o atendimento sempre que for necessário e não deixar de atender nenhum cliente, mesmo que sejam milhões de pessoas, e coletando informações em tempo real sobre as experiências dos usuários pode aprender formas de melhorar continuamente esse processo (ORACLE, 2020).

A respeito desta pesquisa, neste momento não há distinção entre a utilização de um ou outro modelo, pois a ênfase do trabalho é sobre o processo de interlocução entre o usuário e o sistema e como outras tecnologias podem afetar a efetividade deste, sendo comum aos dois tipos, *chatbots* e assistentes virtuais.

3.5 Plataformas Conversacionais

Se verá a seguir alguns exemplos de tecnologias utilizadas como plataformas para a criação, manutenção e hospedagem de *chatbots* nas suas diferentes abordagens, como descritos anteriormente.

Como qualquer sistema mais complexo, essa tecnologia é formada por uma variedade de outras tecnologias conectadas e que permitem o desenvolvimento de *chatbots* mais ou menos integrados, de acordo com as necessidades do usuário. A partir de Medium (2020), se vai enumerar alguns desses módulos que fazem parte de um sistema de *chatbot*.

3.5.1 Plataformas de conversação

São denominadas plataformas de conversação todo e qualquer sistema desenvolvido para servir de meio para que os usuários possam interagir entre si ou com sistemas automatizados. Dessa forma, para disponibilizar informações aos usuários de um sistema específico é necessário escolher o meio que vai permitir isso, sendo considerada essa a decisão mais importante, pois escolhendo uma plataforma em que os usuários já utilizam e tem familiaridade, os problemas de aprendizado de uso da ferramenta podem ser minimizados e a curva de utilização pode ser muito mais rápida. Visto que atualmente todos os aplicativos de mensagens possuem suporte para a tecnologia dos *chatbots*, a decisão acaba ficando mais fácil.

Discord⁴³

O Discord é um aplicativo para comunicação em jogos e em *podcasts*, é focado em *chat* de voz e texto e tem muitos recursos que foram atribuídos para garantir a melhor experiência no uso, incluindo recursos como tempo de resposta muito rápido, servidores de bate-papo de voz gratuitos e a possibilidade de criação de espaços controlados por convite.

Facebook Messenger⁴⁴

É um aplicativo de mensagens que pode ser instalado no celular na forma de um *app*, ser utilizado pela Internet através do seu *site* ou mesmo utilizado dentro da plataforma da rede social do Facebook. O aplicativo está sempre em aperfeiçoamento com novidades para competir com vários outros concorrentes parecidos. A plataforma estimula a utilização dos recursos por empresas para colocar a marca mais próxima dos clientes e oferece uma série de recursos para tal, como a possibilidade do uso de *chatbots* para o atendimento.

Kik⁴⁵

Como alternativa a outros aplicativos de bate-papo, como WhatsApp e Facebook Messenger, o Kik permite falar com outros usuários sem precisar de número de celular e disponibiliza uma grande variedade de *emojis* e filtros e até mesmo uma loja de *bots* para os usuários utilizarem em suas conversas e grupos. Estima-se que a maioria dos usuários da plataforma são adolescentes.

Slack⁴⁶

Consiste numa ferramenta de comunicação e colaboração para equipes de projetos, disponível em todos os tipos de dispositivos para trabalhar com equipes dentro e fora da empresa. Possui mais de 750 mil empresas clientes, grandes e pequenos negócios, e trabalha com um grau bastante elevado de segurança e criptografia.

SMS (*Short Message Service*)

Possibilita a comunicação entre praticamente todos os aparelhos celulares do mundo e pode ser usado de forma automática para interação entre as empresas e os clientes. Existem

⁴³ Disponível em: <https://discord.com>. Acesso em: 21 ago. 2021.

⁴⁴ Disponível em: <https://www.messenger.com>. Acesso em: 21 ago. 2021.

⁴⁵ Disponível em: <https://www.kik.com>. Acesso em: 21 ago. 2021.

⁴⁶ Disponível em: <https://slack.com>. Acesso em: 21 ago. 2021.

ferramentas que podem interagir com as pessoas como um bate-papo usando as mensagens SMS como meio.

Telegram⁴⁷

É um aplicativo de mensagens como seu maior concorrente, o WhatsApp, e fornece um aplicativo com foco em velocidade e segurança, sincronizando suas mensagens em todos os seus dispositivos e permite enviar mensagens, fotos, vídeos e arquivos, bem como criar grupos para até 200.000 pessoas ou canais para transmissão para públicos ilimitados. Atualmente tem um público de mais de 400 milhões de usuários ativos.

Twitter⁴⁸

É uma rede social e um serviço de *microblog* utilizado por milhões de pessoas e empresas pelo mundo. Um dos grandes atrativos da rede é a possibilidade de obter informações sobre o que acontece no mundo em tempo real. A presença de *bots* nesta rede tem sido evidenciado há algum tempo devido a inúmeras acusações que poderiam influenciar as eleições em vários países, devido a quantidade de usuários e a capilaridade de suas notícias. Em julho de 2020, a rede contava com mais de 186 milhões de usuários diários com uma taxa constante de crescimento nos últimos anos.

Website

É possível colocar um *chatbot* dentro de uma página na Internet, de uma empresa por exemplo, e quando um cliente entrar para saber mais seus produtos, pode tirar suas dúvidas pelo *chat* ou, até mesmo, comprar produtos por meio de uma conversa com um sistema de diálogo.

WeChat⁴⁹

Mais de um bilhão de usuários no mundo inteiro utilizam esse aplicativo. Além de *chat* e chamadas de voz e vídeo, ele permite uma série de outros recursos, como jogar, ler notícias, usar serviços locais pelo aplicativo, efetuar pagamentos móveis, entre muitas outras funções.

⁴⁷ Disponível em: <https://telegram.org>. Acesso em: 21 ago. 2021.

⁴⁸ Disponível em: <https://twitter.com>. Acesso em: 21 ago. 2021.

⁴⁹ Disponível em: <https://www.wechat.com>. Acesso em: 21 ago. 2021.

WhatsApp⁵⁰

É um aplicativo de mensagens usado por mais de dois bilhões de pessoas, em mais de 180 países, segundo seu próprio *site*, e foi desenvolvido como alternativa ao sistema de SMS, muito usado no mundo todo, permitindo aos seus usuários o envio e recebimento de diversos arquivos de mídia: textos, fotos, vídeos, documentos e localização, além de chamadas de voz. Promete um sistema de criptografia de ponta a ponta onde ninguém, a não ser os interlocutores, têm acesso ao conteúdo das mensagens.

3.5.2 Assistentes virtuais

Além dos aplicativos de *chat*, existem vários assistentes pessoais de grandes empresas que podem ser instalados em aparelhos celulares e em outros equipamentos, que permitem inclusive controlar uma casa inteligente.

Alexa (Amazon)⁵¹

A Alexa é um serviço na nuvem que você controla com a sua voz e pode fazer inúmeras funções como fazer chamadas telefônicas, controlar uma casa inteligente, ouvir notícias, ouvir música, obter informações sobre praticamente qualquer coisa, controlar a TV, fazer compras no portal da Amazon, executar pequenos programas de terceiros, entre muitas outras funções. Ela pode ser instalada em aparelhos celulares, mas também pode ser comprada na forma de um pequeno aparelho composto de caixa de som e microfone para ser instalado em qualquer ambiente.

Bixby (Samsung)⁵²

Esse aplicativo vem instalado nos celulares de ponta da Samsung e permite fazer uma série de funções, como por exemplo, controlar o celular, efetuar comandos com recurso de voz, controlar dispositivos de casas inteligentes, permite utilizar diversos recursos de acessibilidade por voz usando o celular inclusive como olhos para identificar objetos ou cores, criar rotinas automatizadas de tarefas, entre muitas outras competências.

⁵⁰ Disponível em: <https://www.whatsapp.com>. Acesso em: 21 ago. 2021.

⁵¹ Disponível em: <https://alexa.amazon.com.br>. Acesso em: 21 ago. 2021.

⁵² Disponível em: <https://www.samsung.com/br/apps/bixby>. Acesso em: 21 ago. 2021.

Cortana (Microsoft)⁵³

Cortana é uma assistente virtual inteligente que foi disponibilizada a partir da versão 10 do sistema operacional Windows da Microsoft. A partir dela é possível o controle da agenda, utilização de programas por meio dela, controle de lista, envio de mensagens e participação de conferências, controle de equipamentos inteligentes, atualizações meteorológicas e de tráfego, entre outros.

Google Assistant (Google)⁵⁴

Com este assistente que está instalado em praticamente todo dispositivo Android é possível gerenciar suas atividades, receber ajuda com tarefas cotidianas, controlar dispositivos domésticos inteligentes, se comunicar por meio de mensagens, e-mails e ligações por viva-voz, localização de rotas e ajuda com mapas, pesquisas na *Web* etc.

Oracle Digital Assistant (Oracle)⁵⁵

É um sistema desenvolvido para empresas e pago, mas contempla várias tecnologias de ponta da própria empresa e permite ao cliente utilizá-las ao desenvolver seu assistente. Permite acesso a aplicativos de negócios por meio de *chatbots* e interfaces de inteligência artificial de conversação, disponibilizando para tal um conjunto de ferramentas bastante completo, inclusive com assistentes digitais pré-criados e personalizáveis. Através de uma combinação Inteligência Artificial com análise semântica profunda de PLN, CLN e algoritmos personalizados de aprendizagem de máquina é possível fornecer uma ferramenta sem código para o desenvolvimento de assistentes empresariais com acesso direto às informações gerenciais.

Siri (Apple)⁵⁶

É um assistente instalado nos equipamentos da empresa Apple. Possui grande integração com todo o ecossistema da empresa, através de comandos em linguagem natural, é possível fazer chamadas de voz e vídeo, enviar mensagens, controlar aparelhos de casas conectadas, controlar agendas, consultar trânsito e clima, encontrar e tocar músicas, responder as dúvidas do usuário, controlar outros aplicativos, dentre muitos outros recursos.

⁵³ Disponível em: <https://www.microsoft.com/en-us/cortana>. Acesso em: 21 ago. 2021.

⁵⁴ Disponível em: <https://assistant.google.com>. Acesso em: 21 ago. 2021.

⁵⁵ Disponível em: <https://www.oracle.com/da>. Acesso em: 21 ago. 2021.

⁵⁶ Disponível em: <https://www.apple.com/br/siri>. Acesso em: 21 ago. 2021.

3.5.3 Ferramentas para construção de *chatbots*

Durante o processo de desenvolvimento do programa, um passo importante é determinar qual tecnologia deverá ser usada para a criação das regras. Existem várias ferramentas no mercado, algumas inclusive que não é necessário ser programador para a codificar, já outras podem necessitar de um programador experiente para codificar a lógica e os parâmetros determinados para a criação do *chatbot* ou assistente. Quanto maior o grau de complexidade do assistente, maior será o número de tecnologias empregadas, principalmente o processamento de linguagem natural e a inteligência artificial.

Alexa Skill Builder⁵⁷

A empresa Amazon fornece um conjunto de ferramentas que chamou de “*skills*”, no português habilidades, que permite ao desenvolvedor criar aplicativos que possam ser utilizados por sua assistente digital, a Alexa, de forma que aprenda novas “habilidades”. Os clientes da plataforma têm acesso às tecnologias que permitem fazer perguntas ou solicitações que podem ser aplicadas sobre as mais diversas fontes de dados, como som, vídeo, automação, jogos, entre outros.

BotKit⁵⁸

É uma ferramenta de código aberto com o propósito de construção de *bots* de bate-papo e aplicativos para as principais plataformas de mensagens. A ideia inicial era simplificar o processo de criação para que qualquer pessoa pudesse ter o seu *bot* da forma que desejasse e com o aumento do uso dessa tecnologia e a relevância dessa empresa nesse mercado, a empresa foi adquirida pela Microsoft em novembro de 2018, e desde então passa a fazer parte da estrutura desta empresa somando a outras iniciativas já em funcionamento e disponibilizadas para a comunidade.

BotPress⁵⁹

Um pacote completo de ferramentas altamente customizáveis. Disponibiliza uma das mais poderosas soluções de inteligência artificial conversacional, segundo o desenvolvedor, que pode se integrar com as principais plataformas e tem suporte a vários idiomas. Desenhada para utilização no meio empresarial e de código aberto desde 2017.

⁵⁷ Disponível em: <https://developer.amazon.com/pt-BR/alexa/alexa-skills-kit>. Acesso em: 21 ago. 2021.

⁵⁸ Disponível em: <https://botkit.ai>. Acesso em: 21 ago. 2021.

⁵⁹ Disponível em: <https://botpress.com>. Acesso em: 21 ago. 2021.

DialogFlow⁶⁰

É um produto ofertado pela empresa Google e consiste numa plataforma de processamento de linguagem natural que permite criar interfaces de conversa para *sites*, aplicativos móveis, plataformas de mensagens conhecidas e dispositivos de Internet das Coisas (*Internet of Things* - IoT). Com um escopo de facilitar esse trabalho de design oferece ferramentas para criação de interações ricas entre os usuários e as empresas, extensa variedade de idiomas, aprendizado de máquina e inteligência artificial aplicados aos assistentes com as mesmas tecnologias usadas pela Pesquisa e pelo Assistente do Google. Pode analisar interações em voz e texto e responder de várias maneiras, inclusive texto e voz sintética.

LUIS⁶¹

Um serviço baseado em aprendizado de máquina para construir linguagem natural em aplicativos, *bots* e dispositivos da Internet das Coisas. É um dos serviços disponibilizados pela Microsoft na sua plataforma de serviços cognitivos. Sendo capaz de interpretar os objetivos (intenções) do usuário, até 500 intenções, para transformar comandos em ações inteligentes.

Microsoft Bot Framework⁶²

Um *framework* de desenvolvimento de *chatbots* bastante abrangente que tem a capacidade de falar, ouvir, entender e aprender com seus usuários, com diversas ferramentas que permitem construir *bots* que interagem naturalmente, soluções de segurança e escalabilidade para corporações e uma grande capacidade de integração com outros sistemas para criar desde um simples *bot* de perguntas e respostas a um assistente virtual bastante sofisticado.

QnA Maker⁶³

Com um conceito de criação, treinamento e publicação de *bots* sem escrever uma única linha de código, oferece conjuntos de dados de bate-papo já construídos e pode utilizar como fontes de dados páginas de perguntas frequentes, *sites* de suporte, manuais de produtos, documentos do *SharePoint* ou conteúdo editorial por meio de interfaces intuitivas e em mais de 50 idiomas.

⁶⁰ Disponível em: <https://cloud.google.com/dialogflow>. Acesso em: 21 ago. 2021.

⁶¹ Disponível em: <https://www.luis.ai>. Acesso em: 21 ago. 2021.

⁶² Disponível em: <https://dev.botframework.com>. Acesso em: 21 ago. 2021.

⁶³ Disponível em: <https://www.qnamaker.ai>. Acesso em: 21 ago. 2021.

Rasa Stack⁶⁴

É uma estrutura de aprendizado de máquina desenvolvido para automatizar assistentes virtuais baseados em texto e voz e foi concebido na forma de código aberto. O nível de *bots* criados por sua estrutura são classificados de assistentes contextuais e se consideram um nível acima dos assistentes pessoais, numa classificação de 5 níveis. Investem em pesquisas de ponta em inteligência artificial conversacional, em entendimento de linguagem natural e conexão com diversos sistemas e bases de conhecimento existentes.

Watson⁶⁵

É uma iniciativa da IBM (*International Business Machines*) no mundo dos *bots*, e é conhecida por ter seu produto, o Watson, ter vencido os campeões de um jogo de respostas na tv americana⁶⁶ a alguns anos totalmente sem o auxílio dos desenvolvedores e da Internet. O sistema foi construído não pensando na interação entre humanos, mas entre humanos e assistentes. A sua inteligência artificial conversacional permite a compreensão de texto e som dos mais diferentes aspectos do diálogo humano e pode interpretar as informações e construir respostas mais inteligentes, bem parecidas com um atendimento humano. Permite integração com as maiores plataformas de atendimento ao cliente do mundo e as principais plataformas de *call center*, onde pode atender, por meio de voz, milhares de atendimentos simultâneos.

Wit.ai⁶⁷

Plataforma de processamento de linguagem natural de propriedade do Facebook. Permite a criação de *bots* para comunicação em diversas plataformas, utilização em sistemas de casas inteligentes e dispositivos vestíveis. Possui suporte até o momento a 132 línguas e uma comunidade de mais de 200.000 desenvolvedores.

Woebot⁶⁸

É uma ferramenta baseada em bate-papo que fornece um conjunto de programas de terapia clinicamente validados para tratamento desde ansiedade e depressão generalizadas até condições específicas como depressão pós-parto, depressão de adultos e adolescentes e abuso

⁶⁴ Disponível em: <https://rasa.com>. Acesso em: 21 ago. 2021.

⁶⁵ Disponível em: <https://www.ibm.com/watson>. Acesso em: 21 ago. 2021.

⁶⁶ Disponível em: <https://gizmodo.uol.com.br/computador-da-ibm-vence-de-lavada-dois-cerebros-humanos-em-jogo-de-conhecimentos-gerais/> Acesso em: 21 ago. 2021.

⁶⁷ Disponível em: <https://wit.ai>. Acesso em: 21 ago. 2021.

⁶⁸ Disponível em: <https://woebothealth.com>. Acesso em: 21 ago. 2021.

de substâncias. É a combinação de pesquisas em psicologia com tecnologias mais modernas de inteligência artificial e processamento de linguagem natural. Onde esse aparato pode determinar as melhores perguntas para entender e avaliar o estado emocional do usuário e direcioná-lo ao tratamento.

X.ai⁶⁹

É uma plataforma de *bot* que permite automatizar a criação e controle de agenda por meio da inteligência artificial. O usuário compartilha um *link* com as opções de tempo e o sistema entra em contato com os convidados para agendar as reuniões, em que estes escolhem um horário de acordo com suas preferências, de forma a encontrar horários compatíveis para reuniões com grandes grupos, envio de acompanhamentos e reagendamento de reuniões.

3.5.4 Persistência de Dados

Ao planejar o desenvolvimento de um *bot*, ou mesmo um assistente virtual, é muito importante definir como será a persistência dos dados que será utilizada pelo sistema, bem como a forma como os dados coletados serão analisados para que a solução aprenda e possa refinar cada vez mais os resultados entregues aos usuários. É importante verificar se as ferramentas selecionadas atendem às recomendações para utilização de certos padrões de bancos de dados, tais como: os relacionais ou SQL (*Structured Query Language*) e os não relacionais ou NoSQL (*Not Only SQL*). Alguns exemplos utilizados, são descritos no Quadro 2. Caso não exista a indicação de um sistema de banco de dados específico, podem ser escolhidos de acordo com a familiaridade e disponibilidade do desenvolvedor. A seguir algumas soluções de banco de dados recomendados por algumas plataformas e outras de uso geral que podem ser usadas.

AirTable⁷⁰

Um banco de dados utilizado por mais de 200.000 organizações, mas que possui uma interface bastante amigável parecida com uma simples planilha eletrônica. Permite integração com diversos *sites* e aplicativos e fornece recursos para automação de trabalhos entre a ferramenta e outros aplicativos.

⁶⁹ Disponível em: <https://x.ai>. Acesso em: 21 ago. 2021.

⁷⁰ Disponível em: <https://airtable.com>. Acesso em: 21 ago. 2021.

Datastore⁷¹

Fazendo parte da estrutura do Google, o Cloud Datastore é um banco de dados NoSQL altamente escalonável, ideal para aplicativos da Web e de dispositivos móveis.

Firestore⁷²

Mantido pelo Google, ele permite o desenvolvimento de maneira facilitada de aplicativos complexos usando um banco de dados de documentos totalmente gerenciado, escalonável e sem servidor.

MongoDB⁷³

É um banco de dados orientado a documentos baseado na nuvem, com código aberto, segurança bastante elaborada, com diversas funcionalidades para que seu acesso seja o mais rápido e sem latência (atrasos), inclusive distribuindo ele por diversas regiões e várias nuvens.

Postgress SQL⁷⁴

Um banco de dados relacional com muitos recursos para os desenvolvedores, de código aberto, alta confiabilidade, com mais de 30 anos de desenvolvimento e multiplataforma.

MySQL⁷⁵

É o banco de dados de código aberto mais popular do mundo. O MySQL é um sistema de gerenciamento de banco de dados, que utiliza a linguagem SQL como interface. E atualmente é um dos sistemas de gerenciamento de bancos de dados fornecidos pela Oracle Corporation.

3.5.5 Métricas de Avaliação

Outra etapa importante do processo de desenvolvimento é a análise das métricas de avaliação e informação sobre o funcionamento do sistema de *bot*. Isso é importante para o contínuo aprendizado da plataforma e as devidas intervenções humanas nas regras e decisões que o sistema tomou, pois existem situações que não foram previstas durante o desenvolvimento do *bot* que pode ser analisada posteriormente e mudanças podem ser feitas de modo que não volte a ocorrer. O processo de desenvolvimento de um *chatbot* é um processo constante e deve

⁷¹ Disponível em: <https://cloud.google.com/datastore>. Acesso em: 21 ago. 2021.

⁷² Disponível em: <https://firebase.google.com/products/firestore>. Acesso em: 21 ago. 2021.

⁷³ Disponível em: <https://www.mongodb.com>. Acesso em: 21 ago. 2021.

⁷⁴ Disponível em: <https://www.postgresql.org>. Acesso em: 21 ago. 2021.

⁷⁵ Disponível em: <https://www.mysql.com>. Acesso em: 21 ago. 2021.

ser sempre aprimorado, para isso existem ferramentas que podem ser disponibilizadas pelos módulos citados anteriormente, ou mesmo recomendado pelos desenvolvedores da ferramenta escolhida para a construção do *chatbot*. Existem ferramentas gerais que podem ser aplicadas para esse uso, mas também podem ser utilizadas ferramentas destinadas para esse uso na avaliação das plataformas, as quais são apresentadas a seguir.

Chatbase⁷⁶

A ferramenta Chatbase permite medir, analisar e otimizar o atendimento ao cliente pelos agentes virtuais. Por meio de relatórios interativos é possível identificar o envolvimento do usuário e detectar possíveis problemas de conversação entre ele e o agente. Através da análise do fluxo das conversas é possível acompanhar o desenvolvimento do agente conversacional e melhorar a forma de interação com o cliente. Há pouco tempo foi introduzido na plataforma do Google e agora o Chatbase é o provedor de análise da solução Google Cloud Contact Center AI.

Dashbot⁷⁷

O Dashbot é um sistema para análise de *chatbots* e voz com propósito de otimizar a eficácia da resposta, entender os comportamentos do usuário e aumentar sua satisfação, engajamento e monetização da solução. Algumas funções da ferramenta permitem análises conversacionais de frases, expressões e intenções, tratamento do engajamento e retenção de usuários, determinação do comportamento e intenções do usuário, descobrir a demografia do público, métricas de comparação entre usuários e mensagens, disponibilização de ferramentas para ações de intervenção e acessos multiusuários e multiclientes.

Google Analytics⁷⁸

É a proposta do Google para análise de negócios, que fornece ferramentas para ajudar o administrador a entender seus clientes, melhorar seu modelo de negócio e gerar melhores resultados. Por meio da inteligência artificial e aprendizagem de máquina a solução pode gerar relatórios e painéis detalhados com a situação atual e o histórico do funcionamento do seu negócio para encontrar os possíveis problemas ou gargalos e saná-los. Está disponível para

⁷⁶ Disponível em: <https://chatbase.com>. Acesso em: 21 ago. 2021.

⁷⁷ Disponível em: <https://www.dashbot.io>. Acesso em: 21 ago. 2021.

⁷⁸ Disponível em: <https://analytics.google.com>. Acesso em: 21 ago. 2021.

utilização por todas as soluções disponibilizadas pelo seu ecossistema no Google Cloud, além de ser disponibilizado para ser utilizado por plataformas de terceiros.

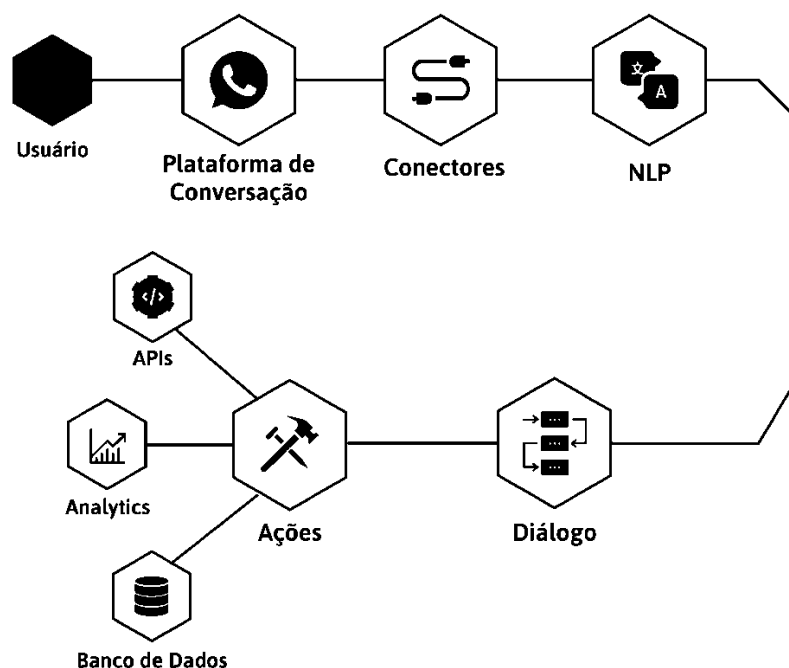
3.6 Módulos de um *Chatbot*

A construção de uma interface conversacional exige uma série de decisões, como demonstrado anteriormente, e neste momento se discutirá algumas tecnologias e estas como se conectam para formar o sistema. Devido a inerente complexidade da construção de tal tecnologia, é necessário selecionar e utilizar uma quantidade grande de ferramentas que estão disponíveis, pois muitas empresas disponibilizam tecnologias para prover essas soluções. Por isso, existem muitas formas de fazer a mesma coisa e na sequência, serão discutidos alguns exemplos e como podem ser implementados.

Ao demonstrar como funciona a conexão de módulos na construção de um *chatbot*, Medium (2020) descreve a dificuldade de escolher uma tecnologia dentre a variedade disponível hoje, mas explica que na criação de *chatbots* altamente customizados e profissionais, faz-se necessário escolher as mais completas e que mais se conectam, e outras vezes na mesma solução, já são disponibilizados todos os módulos necessários para a construção do *chatbot*.

O autor descreve como os módulos se conectam, a partir de uma “conversa”, como se fosse um diálogo entre duas pessoas, pois a partir da requisição do usuário ao chatbot a interação entre eles precisa acontecer de forma natural e automatizada, conforme ilustrado na Figura 6,

Figura 6 - Módulos de uma Interface Conversacional



Fonte: Medium (2020)

Um usuário, ao utilizar uma Plataforma de Conversação, como uma rede social ou canal disponibilizado pela empresa para contato, envia uma mensagem que é direcionada até o servidor do *chatbot* e é recepcionado por um Conector, que precisa ser específico para cada Plataforma de Conversação. Neste momento os dados provindos da rede social, por exemplo, são traduzidos para que possam ser processados, separando o nome do usuário da mensagem enviada. Depois que os dados já foram tratados, são enviados para o módulo de Processamento de Linguagem Natural para que este possa identificar diversas partes como as perguntas, objetivos, contextos, sentido e outros parâmetros que podem estar inseridos na fala. Atualmente várias soluções utilizam a Inteligência Artificial neste módulo para processar a extração dos termos.

O módulo de Diálogo recebe as informações e os termos encontrados e os processa para formular uma resposta condizente com a pergunta do usuário, neste momento, ele constrói uma mensagem de resposta e verifica se é necessário que algum dado seja recuperado de outro local.

Após a requisição ao módulo de Ações, que é responsável por coletar informações ou realizar diversas tarefas, tais como buscas em bancos de dados, APIs (*Application Programming Interface*), ferramentas de avaliação ou qualquer fonte de dados atribuída para

fornecer informações para o *chatbot*, este procura as informações e as devolve para o módulo de Diálogo para que este formule as respostas ao usuário.

Após a conclusão dessa etapa, ocorre o caminho de volta, o módulo de Diálogo as devolve para o módulo de Processamento de Linguagem Natural que cria uma resposta para o usuário usando texto ou voz, de acordo com a requisição inicial do usuário e envia de volta para o Conector, que faz o encapsulamento da mensagem e envia de volta para a Plataforma de Conversação que por fim envia a mensagem para o usuário. Todo esse processo leva alguns milissegundos e pode ser mais ou menos elaborado, de acordo com o escopo do sistema de *chatbot*.

Baseado nessa ideia, algumas plataformas combinam diversas tecnologias e recomendam a sua integração na solução de *chatbot* ou assistente. No Quadro 3, tem-se um exemplo mostrado por Medium (2020), que classifica as tecnologias possíveis de acordo com os módulos descritos anteriormente.

Quadro 3 – As plataformas, módulos e tecnologias que compõem um *chatbot*

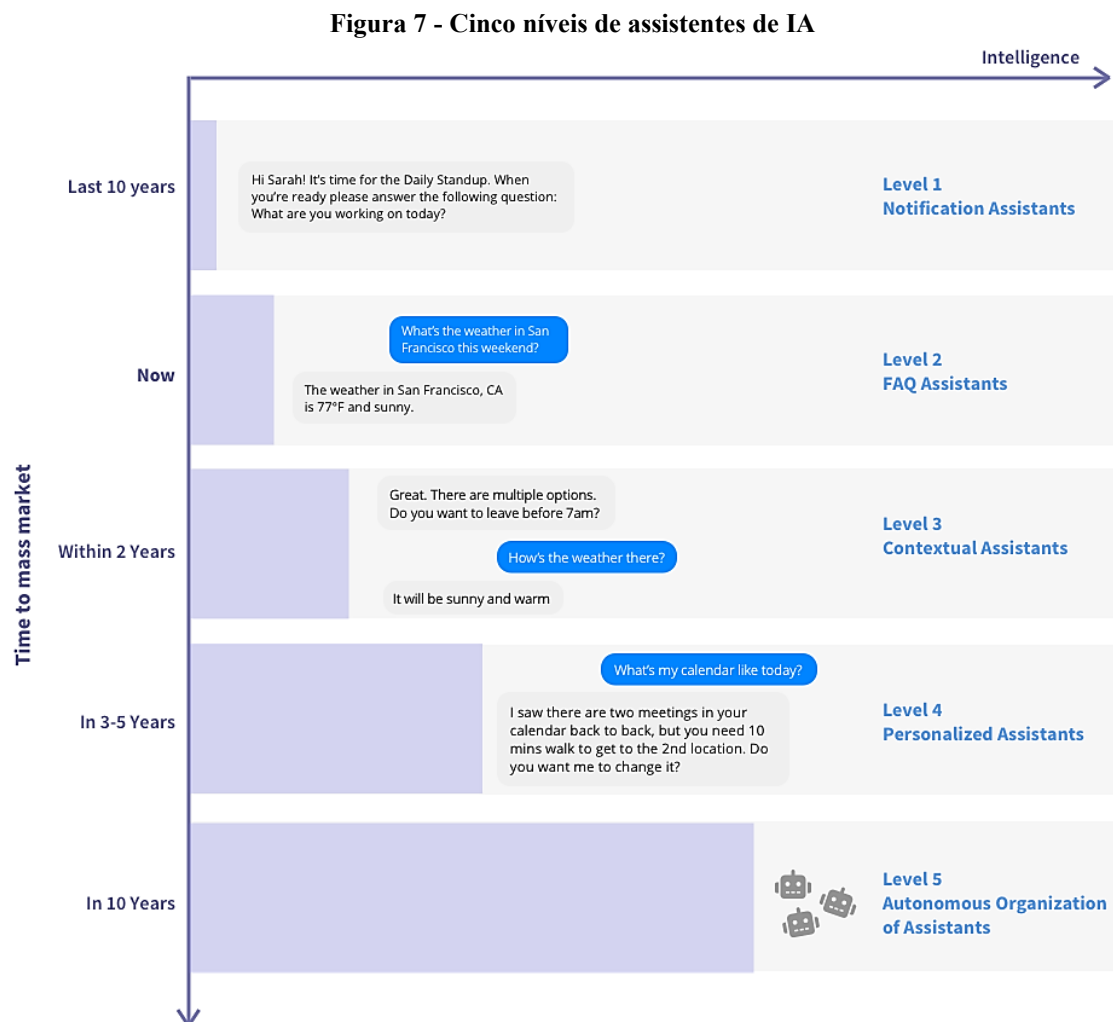
Módulos	Google	Microsoft	BotPress
Diálogo	Dialogflow	BotKit	Botpress
PLN	Dialogflow	LUIS	Botpress
Ações	Firebase Cloud Functions	BotKit + Azure	Botpress + Heroku
Banco de Dados	Firestore	MongoDB (mais fácil) ou bancos relacionais como MySQL e Postgres	PostgreSQL
Analítica	Dialogflow + Firebase	Azure Bot Service	Botpress
Plataformas de Conversação	Integrações diretas Pelo Dialogflow: Google Assistant, Messenger, Telegram, Slack e Twitter.	Cortana, Skype, Microsoft Team, Slack, Messenger, Twilio SMS (Whatsapp e SMS) e Websites	Facebook Messenger, Slack, Microsoft Teams, Skype, Websites e Telegram.

Fonte: Adaptado de Medium (2020)

Assim, a necessidade de planejamento do comportamento do sistema deve influenciar a escolha das ferramentas que serão utilizadas, bem como o conhecimento dos desenvolvedores, visto que até mesmo uma pessoa leiga, com as ferramentas certas, pode conseguir programar seu *bot*, mesmo simples, sem muito conhecimento de programação, mas um sistema complexo pode exigir uma variedade grande de profissionais de diversas áreas para projetar um sistema com respostas cognitivas verdadeiras e que possam interagir de maneira bastante humana com os usuários, e principalmente, entendendo-os e resolvendo seus problemas.

Ao refletir sobre o estado atual e o futuro dessa tecnologia, a possibilidade de classificar os *chatbots* em níveis de Inteligência se torna importante para avaliar o crescimento dessa tecnologia, para isso Nichol (2018) enumerou 5 níveis de inteligência baseado em uma analogia à classificação de carros autônomos, e que no estado atual de desenvolvimento, praticamente quase todos os assistentes estão no segundo nível e está começando a aparecer os primeiros *chatbots* cuja inteligência artificial poderiam ser classificados como terceiro nível.

Ao se dividir em níveis a inteligência artificial dos assistentes, pode-se classificar de simples sistemas de notificação até organizações completamente autônomas, níveis estes mostrados pela Figura 7.



Fonte: (NICHOL, 2018)

Servindo como exemplo, os *chatbots* e assistentes desenvolvidos nos últimos anos, de acordo com Nichol (2018), poderiam ser assim classificados:

Nível 1: Assistentes de Notificação

Sistemas simples de notificações que podem aparecer em seu telefone por meio de um aplicativo de mensagens como o WhatsApp.

Nível 2: Assistentes de Perguntas Frequentes

O tipo de assistente mais comum atualmente, ele permite que o usuário faça uma pergunta e receba uma resposta, quase como uma página de perguntas frequentes, mas a maior diferença é que o *chatbot* tem acesso ao todo e vai entregando apenas o que foi requisitado, às vezes com mais algumas perguntas para o acompanhamento.

Nível 3: Assistentes Contextuais

Um assistente onde o contexto importa muito, pois de acordo com o que o usuário disse antes, quando, onde, como ele disse, e assim por diante, pode alterar o significado e resultar em uma resposta diferente. Para isso o sistema precisa ser capaz de entender o contexto para responder a perguntas diferentes e inesperadas. Começaram a aparecer alguns exemplos, como o Google Duplex⁷⁹.

Nível 4: Assistentes Personalizados

Ao utilizar um assistente desse tipo, ele se lembrará de suas preferências e trabalhará para lhe fornecer uma interface personalizada, pois aprenderá conforme vai lhe conhecendo. Ao conhecer o usuário, poderá ajudá-lo da melhor forma e nos melhores momentos, da mesma maneira como os humanos fazem quando se tornam mais familiares.

Nível 5: Organização Autônoma de Assistentes

Este nível depende de um grande avanço que poderá levar muitos anos, mas é perfeitamente factível e uma realidade. Um grupo de assistentes inteligentes executa grande parte das operações da empresa devido ao fato de conhecer todos os usuários pessoalmente.

A solução Google Duplex que é considerado o primeiro assistente de nível 3, foi apresentada em maio de 2018 como um sistema de inteligência artificial para automação de

⁷⁹ Disponível em: <https://ai.googleblog.com/2018/05/duplex-ai-system-for-natural-conversation.html>. Acesso em: 21 ago. 2021.

interações conversacionais. Ele a partir do pedido do usuário poderia marcar uma consulta médica, um horário no cabeleireiro, uma reserva em um restaurante, completamente sozinho, por meio de voz e uma ligação telefônica para o telefone do estabelecimento (LEVIATHAN; MATIAS, 2018).

A Figura 8 mostra como interage com o assistente dos celulares Android, o Google Assistant, para que a partir de um sistema telefônico de voz, possa contactar outras empresas e efetuar uma transação de dados, mesmo que por meio de um atendente humano, ele pode agir sobre a agenda do estabelecimento, por exemplo, e depois de confirmado ele pode aprovar na agenda do usuário que requereu o contato.

Figura 8 – Google Duplex como módulo do Assistente



Fonte: (LEVIATHAN; MATIAS, 2018)

Com o direcionamento para resolver tarefas específicas, como agendamento de determinados tipos de consultas, o sistema pode explorar extensivamente esse domínio para aprender todas as nuances da fala humana e conseguir ser tão natural. Essa naturalidade é devido a tecnologia por trás da solução. Seu núcleo é formado por uma rede neural recorrente projetada para esse fim, que permitiu os avanços necessários na compreensão, interação, tempo e fala. Além disso, essa rede precisou de muito treinamento, e foi utilizado um corpus de dados de conversação telefônica anonimizados.

Assim, pode se dizer que a tecnologia dos assistentes digitais, *chatbots* ou apenas *bots* está em alta no setor tecnológico e chegando cada dia a mais empresas e usuários, com a função de ajudar e reduzir custos. Muita pesquisa está sendo desenvolvida pelas maiores empresas dessa área para alcançar níveis nunca atingidos de interação com as máquinas e novas tecnologias que estarão disponíveis para trabalhar em diversas áreas, oriundas dessa demanda.

4.

METADADOS

Com o crescimento do volume de dados e sua disponibilização, em grande parte, pela Internet, foi necessário o desenvolvimento de mecanismos para o tratamento adequado dessas informações, tanto para a criação, mas também para a preservação, uso, acesso e recuperação. No passado, o problema de encontrar uma informação em um repositório era basicamente resolvido dentro das bibliotecas, mas com a chegada da Internet a todos os lugares e sua disseminação foi possível utilizá-la como principal meio de busca, e consequentemente os problemas de acesso passaram a ser comuns a todas as áreas.

A Ciência da Informação pôde contribuir como uma área de conhecimento, uma ciência que desde sua gênese participa de todos os processos informacionais, definida por Borko (1968, p. 3, tradução nossa) como,

[...] uma ciência interdisciplinar que investiga as propriedades e o comportamento da informação, as forças que governam os fluxos e o uso da informação, e as técnicas tanto manual quanto mecânica de Processos de Informação para otimizar o armazenamento, a recuperação, e a disseminação da informação.⁸⁰

Assim, com a contribuição destes profissionais no desenvolvimento de técnicas de indexação das informações, e seu entendimento a respeito do tratamento da informação para investigação da eficiência na busca, acesso ou recuperação; permitiu influenciar a forma como é feita a descrição dos dados. Um dos meios de representação das características de uma informação é denominado metadado.

⁸⁰ [...] is that discipline that investigates the properties and behavior of information, the forces governing the flow of information, and the means of processing information for optimum accessibility and usability. It is concerned with that body of knowledge relating to the origination, collection, organization, storage, retrieval, interpretation, transmission, transformation, and utilization of information.

4.1 Definição de Metadados

O termo *metadata* (metadado) foi cunhado na década de 1960 por Jack E. Myers, presidente e fundador da *The Metadata Corporation*, e tinha a proposta de diferentes aplicações nas gestões de recursos informacionais para descrever conjuntos de dados que poderiam ser utilizados na organização, na representação e na localização de recursos (SMIRAGLIA, 2005).

Baseado neste cenário informacional, o fato de fornecer meios para que uma informação seja encontrada é muito relevante, como disse Marcondes (2001), “... a informação relevante para um dado problema tem que estar disponível no tempo certo. De nada adianta a informação existir se quem dela necessita não sabe da sua existência ou se ela não puder ser encontrada”.

O ato de produzir e padronizar os metadados já é descrito há séculos, feito por bibliotecários e indexadores (MILSTEAD; FELDMAN, 1999). Os metadados fornecem recursos úteis como uma forma de auxiliar este processo de encontrar os dados, enriquecendo-os com mais informações. Frequentemente os metadados são descritos como “dados sobre dados”, definição esta que tem sua origem na ciência da computação, ou outra como, “dados descrevendo outros dados”, especialmente utilizado no meio digital, mas é necessário o devido contexto, pois cada vez mais envolvem outras tecnologias e são aplicados de muitas maneiras nas mídias digitais e em outras áreas.

Para entender os metadados em um sentido geral, é importante ter em mente alguns pontos:

- metadados são informações ou dados associados a outros recursos de informação,
- metadados são informações estruturadas,
- metadados são usados para habilitar uma variedade de funções em relação ao recurso que descreve.

Ao se referir a um dado como um recurso, é possível descrever diversos elementos a respeito de suas características, os metadados, e que segundo Alves (2010, p. 47), podem ser descritos como,

[...] atributos que representam uma entidade (objeto do mundo real) em um sistema de informação. Em outras palavras, são elementos descritivos ou atributos referenciais codificados que representam características próprias ou atribuídas às entidades; são ainda dados que descrevem outros dados em um sistema de informação, com o intuito de identificar de forma única uma entidade (recurso informacional) para posterior recuperação.

Na conceituação do metadados digitais, Vellucci (1998, p. 192), destaca a importância de conceituar dado e informação valorizando o contexto e o usuário, de forma a esclarecer que metadados não são dados catalográficos ou bibliográficos com outro nome, além disso os descreve como um

(...) dado que descreve atributos de um recurso, caracteriza suas relações, apoia sua descoberta e uso efetivo, e existe em um ambiente eletrônico. Usualmente consiste em um conjunto de elementos, cada qual descrevendo um atributo do recurso, seu gerenciamento, ou uso.

Ainda sobre os metadados digitais, outra definição é sobre os metadados serem informações compreensíveis por máquina a partir de recursos da Web. Essa ideia de "compreensível por máquina" é muito importante, pois tendo uma semântica e uma estrutura bem definida, isso permite que os agentes de *software* possam usar isso para garantir que os princípios sejam obedecidos, uma forma de garantir a confiança no que está sendo feito e permitindo que todo o processo possa funcionar de uma maneira mais suave e rápida. Para citar um exemplo de uso de metadado na Web, quando um usuário digita um endereço de um sítio na Internet em seu navegador, um objeto usando o protocolo HTTP (*Hypertext Transfer Protocol*) é recuperado, esse protocolo permite que várias outras informações sejam enviadas ao usuário, como informações sobre a sua data, sua data de validade, o proprietário dos recursos e muitas outras informações, que são enviadas pelo servidor ao navegador do usuário. Portanto a *World Wide Web* é um mundo repleto de informações e muitas dessas informações são informações sobre informação (BERNERS-LEE, 1997).

Com o objetivo de identificar as várias funcionalidades dos metadados, De Barros Campos (2007), enumerou dez temas, descrevendo-os como categorias funcionais dos metadados, que tem por objetivo ressaltar as diferentes ênfases em sua utilização, assim nomeadas:

- **Descritividade:** É a principal função de um metadado, que é descrever o recurso ou objeto informacional de modo a permitir sua identificação, localização, recuperação, manipulação e uso. Um documento, no meio digital, é normalmente descrito e estruturado por meio de uma metalinguagem de marcação, e como exemplo de linguagens utilizadas tem-se o HTML (*HyperText Markup Language*) e o XML (*Extensible Markup Language*).
- **Padronização:** É o resultado da aceitação de convenções como compromisso para o entendimento das representações, de forma que a descrição seja uniforme

e todos utilizem da mesma maneira, como as estruturas, convenções, sintaxe e significado consensual dos metadados.

- **Interoperabilidade:** Em um ambiente padronizado de descrição é possível transferir e utilizar informações entre sistemas com eficiência e uniformidade, isso é a definição de interoperabilidade. Como exemplos de linguagens de padronização e interoperabilidade são os padrões de metadados utilizados para a catalogação MARC (*Machine Readable Cataloging*) e Dublin Core (*Dublin Metadata Core Element Set*).
- **Modularidade:** Um objeto digital não deve ser visto e representado como um bloco, pois devido a suas diversas características ele pode ser mais bem compreendido em dimensões ou módulos. Dessa forma, é melhor que a escolha de metadados sejam definidos a partir dos módulos, direcionando para cada módulo os metadados específicos, pode resultar na melhora da praticidade. Um padrão modularizado para a construção de metadados é o RDF (*Resource Description Framework*), por exemplo. E para a construção em camadas de metadados com vistas a uma semântica o DTD (*Document Type Definition*).
- **Reflexibilidade:** Pode ser descrita como a funcionalidade de revelar sua estrutura e padrões utilizados, ou seja, em um ambiente de Internet é necessário apresentar não apenas o conteúdo do objeto informacional, mas também a caracterização, descrição e representação dos próprios sistemas, regras e processos que representam, organizam e tratam esses objetos.
- **Visibilidade:** No momento em que, além do objeto informacional, diversas informações são fornecidas para auxiliar na utilização dos dados, a visibilidade acaba sendo aprimorada devido a essa situação. Também está relacionada à documentação, confiabilidade, facilidade de manutenção, interoperabilidade e potencialidade de usos futuros.
- **Flexibilidade:** Esta funcionalidade é descrita como o modo de lidar com a diversidade de inúmeras estruturas e padrões e permitir que seja possível a interoperabilidade e a automatização dos processos. Devido ao fato de que uma padronização total não é viável, a flexibilidade do sistema é o que permite um bom nível de interoperabilidade.
- **Automatização:** É quando o processo pode ser realizado com o mínimo de intervenção humana, ou até mesmo, quando ao invés da presença humana para

construção de sistemas automáticos, podem ser utilizados as ontologias a fim de obter sistemas flexíveis e que podem atuar em nível semântico e relacionados aos metadados.

- **Administração:** Metadados destinados à administração são constantemente utilizados em repositórios, os metadados não apenas descrevem e identificam um objeto informacional, mas especificam as condições corretas ou ideais de seu gerenciamento, as relações do objeto com outros na coleção, sua função, utilização, comportamento, contexto de criação e condições de preservação. METS (*Metadata Encoding and Transmission Standard*) é uma estrutura de metadados para manutenção e troca de objetos digitais.
- **Preservação:** Preservação e Administração estão bastante ligadas, pois ambos tratam da manutenção e preservação dos dados dos objetos. OAIS (*Open Archival Information System*) é um modelo de referência internacional para metadados de preservação, e que inclui um esquema de informação para inclusão dos metadados aspirando a preservação de informação digital por longo prazo.

Nesse ambiente da Web, em que existe a necessidade de se identificar e recuperar grande volume de informações, os metadados hoje são bastante utilizados.

Um dos maiores objetivos do uso de metadados no contexto da Web é permitir não só descrever documentos eletrônicos e informações em geral, possibilitando sua avaliação de relevância por utilizadores, mas também permitir agenciar computadores e programas especiais, robôs e agentes de “software”, para que eles compreendam os metadados associados a documentos e possam então recuperá-los, avaliar sua relevância e manipulá-los com mais eficiência. (CORREIA; 2017, p. 94)

O fornecimento de metadados é um requisito fundamental, como descrito pela W3C (2017), se não houver metadados disponíveis ou mesmo de qualidade, os dados disponibilizados na Web podem não ser encontrados, e com isso acabam não sendo reutilizados, ou não acessados por outros usuários além do criador. Além disso, complementa, que

Os metadados podem ser utilizados para auxiliar a realização de tarefas tais como a descoberta e reutilização de conjuntos de dados, e podem ser designados de forma a considerar diferentes níveis de granularidade: desde uma propriedade singular de um recurso, até de um conjunto de dados completo ou até mesmo de todos os conjuntos de dados de uma organização específica. Os metadados também podem ser de diversos tipos. Estes tipos podem ser classificados em taxonomias diferentes, seguindo diferentes critérios de agrupamento. Por exemplo, uma taxonomia específica poderia definir três tipos de metadados de acordo com características descritivas, estruturais e administrativas. Uma taxonomia diferente, por sua vez, poderia definir tipos de metadados de acordo com as tarefas nas quais os metadados são utilizados, como por exemplo, a descoberta e reutilização (W3C, 2017).

Ainda prossegue, para a utilização dos metadados em aplicativos Web, recomenda entre outras coisas, o fornecimento de metadados tanto para usuários humanos quanto para os computadores, porque ao fornecer tais informações pode auxiliar as pessoas e aplicações de computadores que não se conhecem mutuamente a descreverem um mesmo conjunto de dados e poderem compreendê-los e reutilizá-los. Fornecer metadados que descrevam as características gerais dos conjuntos de dados e das distribuições, para que possibilite aos agentes de aplicativos descobrir automaticamente os conjuntos de dados que estão disponíveis na Web. E fornecer metadados que descrevam o esquema e a estrutura interna de uma distribuição para que todos que desejam explorar os dados disponibilizados neste conjunto, e auxiliando aos usuários compreender o significado dos dados e permitir os reutilizá-los.

Ainda sobre a facilidade de encontrar um dado por meio de um metadado, Marcondes (2001) complementa que o uso descentralizado de esquemas de metadados com o intuito de descrever e indexar informações disponibilizadas na Web quando são publicados, embutindo os metadados nos códigos das páginas, pode facilitar aos robôs, ou “aranhas”, dos buscadores de conteúdo, a identificação das páginas com informações mais relevantes.

4.2 Padrões de Metadados

A utilização de padrões de metadados por diversas instituições permite a troca de informações entre elas de forma que não haja o retrabalho de descrição dos dados e dos repositórios. A escolha adequada de um padrão permite a interoperabilidade entre sistemas com maior eficiência e rapidez, para isso é necessária dedicação na estimativa das necessidades da instituição e principalmente nas necessidades dos usuários.

Um padrão de metadado possui certas características, e que segundo Alves (2010, p. 47-48) são:

[...] estruturas de descrição constituídas por um conjunto predeterminado de metadados (atributos codificados ou identificadores de uma entidade) metodologicamente construídos e padronizados. O objetivo do padrão de metadados é escrever uma entidade gerando uma representação unívoca e padronizada que possa ser utilizada para recuperação da mesma.

Sobre os diferentes tipos de padrões encontrados, Alves e Santos (2013, p. 65), enfatizam que os metadados estão presentes em muitos sistemas informacionais, e para isso eles precisam ser construídos em diversos tipos, estruturas, níveis de granularidade e especificidade, e os classificam da seguinte forma:

- Metadados para propósitos gerais
 - Nível 1 – Padrões de metadados simples
 - Nível 2 – Padrões de metadados estruturados
- Metadados para propósitos específicos
 - Nível 3 – Padrões de metadados ricos

Classificados de acordo com o nível de suas estruturas de descrição, são divididos em 3 níveis, chamados de formatos de metadados, descritos da seguinte forma:

Nível 1 – Padrões de metadados simples: constituídos por metadados não estruturados, em alguns casos extraídos de forma automática por robôs, apresentam na maioria das vezes uma semântica reduzida. Ex.: *MetaTag(s)* e metadados utilizados na transferência de dados por meio do protocolo HTTP – *Hypertext Transfer Protocol*.

Nível 2 – Padrões de metadados estruturados: constituídos por metadados mais estruturados e definidos em um esquema de metadados, são baseados em normas emergentes e que proporcionam uma descrição mínima do recurso para a sua identificação, localização e recuperação. A descrição geralmente é feita em campos e nessa categoria começa a ser inserida a ajuda de especialistas em informação. Ex.: padrão de dados Dublin Core - DC.

Nível 3 - Padrões de metadados ricos: também considerados padrões de metadados altamente estruturados, são constituídos por metadados complexos e apresenta uma estrutura de descrição mais formal e detalhada. São baseados em normas e códigos especializados de um domínio particular, possibilitam a descrição de um recurso informacional individual ou pertencente a uma coleção e facilitam a localização, a recuperação e o intercâmbio dos recursos informacionais. Ex.: padrão de metadados ou formato MARC 21, da área de Biblioteconomia (ALVES; SANTOS, 2013, p. 65)

Como exemplos de padrões tem-se o MARC e o Dublin Core. O padrão MARC ou catalogação legível por computador, foi desenvolvido na década de 60 pela LC - *Library of Congress*, para possibilitar a catalogação e intercâmbio de registros bibliográficos legíveis por máquina. Legível por computador significa que um tipo específico de máquina, um computador, pode ler e interpretar os dados no registro de catalogação. Catalogar um registro significa criar um registro bibliográfico, ou determinar as informações tradicionalmente mostradas em um cartão de catálogo. O registro inclui, não necessariamente nesta ordem, (1) uma descrição do item, (2) entrada principal e entradas adicionadas, (3) títulos de assunto e (4) a classificação ou número de chamada.

As informações de um cartão de catálogo não podem simplesmente ser digitadas em um computador para produzir um catálogo automatizado. O computador precisa de um meio de interpretar as informações encontradas em um registro de catalogação. O registro MARC contém um guia de seus dados, ou pequenas "placas de sinalização", antes de cada pedaço de informação bibliográfica. O uso do padrão MARC também permite que as bibliotecas façam

uso de sistemas de automação de bibliotecas disponíveis comercialmente para gerenciar as operações da biblioteca. Muitos sistemas estão disponíveis para bibliotecas de todos os tamanhos e são projetados para trabalhar com o formato MARC. A Figura 9 mostra um exemplo de um catálogo feito utilizando MARC. O formato bibliográfico foi atualizado e hoje é chamado de MARC 21, bem como toda a documentação oficial é mantida pela Biblioteca do Congresso dos Estados Unidos e publicado como *MARC 21 Format for Bibliographic Data* (FURRIE, 2009).

Figura 9 – Exemplo de catálogo MARC

"SIGNPOSTS"	DATA
100 1# \$a	Arnosky, Jim.
245 10 \$a	Raccoons and ripe corn /
\$c	Jim Arnosky.
250 ## \$a	1st ed.
260 ## \$a	New York :
\$b	Lothrop, Lee & Shepard Books,
\$c	c1987.
300 ## \$a	25 p. :
\$b	col. ill. ;
\$c	26 cm.
520 ## \$a	Hungry raccoons feast at night in a field of ripe corn.
650 #1 \$a	Raccoons.
900 ## \$a	599.74 ARN
901 ## \$a	8009
903 ## \$a	\$15.00

Fonte: (FURRIE, 2009)

O *Dublin Core Metadata Element Set* é um vocabulário de quinze propriedades para uso na descrição dos recursos que visa descrever objetos digitais, tais como, vídeos, sons, imagens, textos e *sites* na Web. O nome "*Dublin*" deve-se à sua origem em um workshop de 1995 em Dublin, Ohio; e "*Core*" porque seus elementos são amplos e genéricos, utilizáveis para descrever uma ampla gama de recursos. Originalmente eram treze elementos, mas que foi atualizado mais tarde para quinze. Em 1998, isso foi formalizado no padrão RFC 5791 da *Internet Engineering Task Force*, e começaram as discussões sobre torná-la um padrão da Organização Nacional de Padrões de Informação (NISO) (DCMI, 2021).

Esse padrão é mantido pela OCLC (*Online Computer Library Center*) e foi desenvolvido por meio de cooperação internacional. Com a finalidade de permitir uma descrição mais rápida e barata de registros bibliográficos, e para isso foi atribuído algumas características básicas como simplicidade, interoperabilidade semântica, consenso internacional, flexibilidade e modularidade na Web. Por possuir apenas 15 elementos e diversos

qualificadores, não foi projetado com a intenção de substituir modelos mais complexos como a catalogação AACR2 (*Anglo-American Cataloguing Rules*) /MARC, mas servir de recurso para profissionais ou leigos para descrição de recursos simples por meio de um conjunto básico de elementos (DE BARROS CAMPOS, 2007).

4.3 Uso de Metadados

Antes de discutir algumas aplicações dos metadados em domínios digitais e exemplos no escopo dessa pesquisa, pode-se contextualizar algumas de suas funções primárias, como apresentado por Baca (2008, p.13):

- **Criação, multiversão, reutilização e recontextualização de objetos de informação.** Os objetos entram em um sistema de informação digital quando são criados digitalmente ou convertidos em formato digital. Múltiplas versões de um mesmo objeto podem ser criadas para fins de preservação, pesquisa, exposição, divulgação ou mesmo desenvolvimento de produtos. Alguns metadados administrativos e descritivos podem e devem ser incluídos pelo criador ou digitalizador, especialmente se a reutilização estiver prevista, como em um sistema de gerenciamento de ativos digitais.
- **Organização e descrição.** Uma função primária dos metadados é a descrição e ordenação de objetos ou itens originais em um repositório ou coleção, bem como dos objetos de informação relacionados aos originais. Os objetos de informação são organizados automática ou manualmente na estrutura do sistema de informação digital e podem incluir descrições geradas pelo criador original. Metadados adicionais podem ser criados por profissionais da informação por meio de processos de registro, catalogação e indexação ou por outros meios, como folksonomias ou metadados provenientes da contribuição de outros usuários.
- **Validação.** Os usuários examinam metadados e outros aspectos de recursos recuperados, a fim de verificar a autoridade e confiabilidade desses recursos.
- **Pesquisa e recuperação.** Bons metadados descritivos são essenciais para a capacidade dos usuários de encontrar e recuperar metadados e objetos de informação relevantes. Objetos de informação armazenados localmente e virtualmente distribuídos estão sujeitos a busca e recuperação pelos usuários, e

os sistemas de informação criam e mantêm metadados que rastreiam algoritmos de recuperação, transações de usuários e eficácia do sistema em armazenamento e recuperação.

- **Utilização e preservação.** No mundo digital, um objeto informacional pode estar sujeito a muitos tipos diferentes de usos ao longo de sua vida, durante os quais, os seus processos também podem ser reproduzidos e modificados. Metadados relacionados a anotações do usuário, rastreamento de direitos e controle de versão podem ser criados. Os objetos digitais, especialmente os que nascem digitais, também precisam estar sujeitos a um regime de preservação contínuo e passar por processos como atualização, migração e verificação de integridade para garantir sua disponibilidade contínua e documentar quaisquer alterações que possam ter ocorrido no objeto de informação durante os processos de preservação.
- **Disposição.** Os metadados são um componente chave para documentar a disposição de objetos e itens originais em um repositório, bem como dos objetos de informação relacionados a esses originais. Objetos de informação que estão inativos ou não são mais necessários podem ser descartados.

Pensando na disponibilidade das informações em um sítio na Web, com o propósito de organização dos dados e facilidade em sua descoberta e recuperação por meio de um buscador, como o Google, este criou uma categoria de metadados que são impulsionados a serem adicionados no código das páginas Web para esse propósito. Em 2009, o Google introduziu os *rich snippets* nos resultados da busca, algo que pode ser entendido como fragmentos ricos, pois contém informações muito importantes do ponto de vista semântico da página (CASAROTTO, 2017). Um exemplo disso são as avaliações dos usuários, como mostrado na Figura 10.

Figura 10 – Exemplo de *rich snippets*



Rock Content - Página inicial | Facebook
<https://pt-br.facebook.com/werockcontent/> ▼
★★★★★ Avaliação: 4,8 - 475 votos
Rock Content, Belo Horizonte. 88.554 curtidas · 1.254 falando sobre isso · 1.654 estiveram aqui. A Rock Content é Marketing de Conteúdo. Ensinamos,...

Fonte: Casarotto (2017)

Para um usuário que está navegando pela página, por meio dessas informações ele tem acesso a um conteúdo adicional sobre o que está vendo e pode ajudá-lo a decidir qual resultado

pode lhe atender melhor antes mesmo de clicar em qualquer um deles. Já para o lado do buscador, essas informações ajudam a identificar informações relevantes para oferecer resultados mais satisfatórios aos usuários.

Para que o sistema do buscador possa se beneficiar desse material, dados estruturados precisam ser adicionados pelo criador da página e os identificando no código HTML das páginas. O algoritmo do Google já é capaz de capturar todos os códigos das páginas, mas como estes dados são estruturados e padronizados, esse trabalho é mais eficiente e garante maior visibilidade à página que fornece essas informações. O próprio Google fornece informações bastante detalhadas a respeito desses dados para facilitar aos desenvolvedores a codificação desses metadados das páginas, além de deixar bastante claro que a simples existência desse material nas páginas não faz com que elas sejam mais bem ranqueadas, o processo é muito mais complexo, mas a existência desses dados pode contribuir de outras formas para trazer mais público, aumentando a retenção na página e diminuindo a taxa de rejeição dos usuários.

Quando um usuário está procurando alguma informação por meio do buscador Google, ele obtém como resposta diversos resultados que contêm apenas um título, URL (*Uniform Resource Locator*) e uma breve descrição, mas quando os *rich snippets* são encontrados, eles aparecem de forma mais visual e podem atrair o olhar do usuário com mais facilidade, conforme descrito por Casarotto (2017). Existem diversos tipos de *rich snippets*, alguns deles funcionam com qualquer tipo de página e outros são desenvolvidos para tipos de páginas específicas, se vai ver alguns exemplos:

- **Breadcrumbs** - significa “migalhas de pão”, eles descrevem o caminho dentro do *site* para chegar até a página em questão. Eles mostram a sua posição dentro da hierarquia do *site*, em categorias e subcategorias.
- **Sitelinks** - apresentam as principais páginas internas de um *site*, apenas para a primeira página mostrada nos resultados. Ele é construído de maneira automática pelo sistema do buscador, mas se tiver a estrutura das páginas em *rich snippets*, será apresentado de maneira mais coerente e compreensível.
- **Searchbox** – O sistema do Google pode mostrar uma caixa de busca para serem efetuadas pesquisas dentro do *site*, isso acontece quando o sítio tem um tamanho considerável e tráfego significativo e quando é pesquisado pelo usuário o nome de uma marca ou empresa na sua busca primária.

- **Avaliações** – Essas avaliações são exibidas em forma de estrelas, numa escala, e demonstram a opinião dos visitantes sobre o conteúdo que essa página oferece.
- **Logo** – Ao fazer uma busca com o nome de uma marca, empresa ou *site* relevante, o Google pode exibir dados mais específicos sobre eles em uma caixa à direita, chamada de *Knowledge Graph*. Uma das informações apresentadas nessa área é o logo da empresa, que pode estar definida no código e faz com que o sistema utilize a imagem correta ao se tratar dessa empresa.
- **Redes sociais** - Os *links* de redes sociais também podem aparecer no *Knowledge Graph* da sua marca. Eles precisam estar identificados e serão mostrados corretamente.
- **Contato** – Informação importante de uma empresa, ou serviço, são os seus telefones, que pode ser o telefone de contato, tele vendas, Serviço de Atendimento ao Consumidor (SAC), suporte técnico, reserva, entre outros serviços disponibilizados ao consumidor.
- **Negócios locais** – Ao inserir dados estruturados no seu *site*, com informações de endereço, contato, horário de funcionamento e fotos, o sistema ajuda a localizar as empresas próximas e dá destaque a elas além de acrescentar mais informações automáticas como horários de pico, notas e comentários de consumidores e serviços de delivery.
- **Artigos** – Se o sítio da Web possui páginas a respeito de notícias ou posts de blog, ao utilizar dados estruturados eles podem aparecer em forma de um carrossel nos resultados e entre as principais notícias.
- **Produtos** – Se forem adicionadas as informações nos produtos das páginas do seu sítio, o buscador poderá apresentar informações importantes a respeito dos produtos para os usuários como marca, preço, disponibilidade e avaliações.
- **Eventos** – Para páginas de sítios de eventos, a disponibilidade de dados estruturados sobre tais eventos permitem que estes apareçam em destaque nas buscas, incluindo informações como o nome do evento, data, localização, preço de ingressos, entre outros.
- **Receitas** – Uma página de receitas pode apresentar diretamente ao usuário informações como tempo de preparo, calorias, dados nutricionais, avaliações e fotos.

- **Músicas** – Páginas de bandas e cantores ganham destaque e pode além de apresentar os dados da música ou álbum, também agregar os *links* de tocadores de música e vídeo para o usuário.
- **TV e filmes** – Páginas de programas de tv, filmes ou séries recebem destaque pelo buscador ao apresentar ao usuário, pela própria página dos resultados, outras informações como elenco, duração, trailer, premiações e *links* para *streaming* (CASAROTTO, 2017).

Isso é apenas um exemplo da utilização de metadados por uma grande empresa com a pretensão de facilitar o seu próprio trabalho de indexação das páginas na Internet, como ela criou a tecnologia e faz uma campanha bastante didática para influenciar os desenvolvedores e com isso recupera melhores informações e pode apresentar ao usuário que está pesquisando sobre algum assunto melhores resultados e dessa forma consolidando seu lugar como o maior buscador da Internet, além disso, alimenta diversas outras ferramentas e serviços disponibilizados com esses dados enriquecidos que foram coletados.

Como exemplos de metadados aplicados em páginas Web e disponibilizados aos usuários, tem-se o catálogo ATHENA da Biblioteca da UNESP, o BENANCIB e o da Biblioteca Digital Brasileira de Teses e Dissertações (BDTD).

O catálogo disponibilizado pelas bibliotecas da UNESP que é acessado por meio do sistema ATHENA⁸¹ permite acesso a textos disponíveis *online*, como artigos de revistas científicas, livros digitais (*e-book*), dissertações, teses, resenhas, entre outros.

Ao utilizar a plataforma para pesquisar algum texto, o usuário tem acesso a diversas informações na página de resposta, como mostrado na Figura 11.

⁸¹ Disponível em: <http://athena.biblioteca.unesp.br>. Acesso em: 07 jan. 2022.

Figura 11 – Página de busca do catálogo Athena



Fonte: elaborado pelo autor

O usuário pode ter acesso ao documento, quando estiver disponível para download, ou verificar em qual acervo da UNESP está disponível para fazer a requisição ou reservar para buscá-lo em outra data. Além disso, existem diversas formas de exportação dessas informações, dados disponibilizados em forma de metadados que podem ser utilizados para importação por meio de outras ferramentas com todas as informações desta obra. Para verificar o conteúdo de um documento recuperado, tem-se o Quadro 4, com os metadados disponibilizados por meio da exportação em formato Excel (xlsx).

Alguns documentos possuem uma riqueza maior de metadados que outros, mas vários campos são obrigatórios e permitem ter acesso a uma quantidade importante de dados sobre a obra pesquisada que fornecem informações para o usuário tomar a decisão sobre a relevância dos documentos encontrados.

Quadro 4 – Metadados exportados pela página da ATHENA

ATHENA - Biblioteca UNESP Marília	
Título	Novos desafios da educação a distância: programação e uso de <i>Chatbots</i>
Autor	Vieira Barros, Daniela Melaré undefined Guerreiro, Aníbal Martins
É parte de	Revista Espaço Pedagógico, 2019-05-10, Vol.26 (2), p.410-431
Dissertação	
Assunto	EaD, Programação, Tutores Virtuais, Inteligência-Artificial, <i>Chatbot</i> , Novos desafios
Assuntos MESH	
Gênero	

Descrição	A evolução tecnológica, particularmente a Inteligência Artificial, tem contribuído significativamente para mudanças relevantes em quase todas as dimensões da nossa sociedade, nomeadamente na educação e, conseqüentemente, no Ensino à Distância. Tais mudanças exigem transformações significativas, especialmente no processo de ensino-aprendizagem, de forma a minimizar a incapacidade de os Tutores Virtuais responderem na totalidade às solicitações de suporte por parte dos alunos e, por isso, a necessidade enorme de potenciar as tecnologias existentes, com o objetivo de minorar essa dificuldade. Este artigo tem como objetivo compreender os novos desafios da Educação à Distância (EaD) no ensino superior online, nomeadamente a programação e uso de <i>Chatbots</i> , bem como as razões e causas para sua operacionalidade. A metodologia utilizada foi baseada no método dedutivo, resultante do processo de pesquisa bibliográfica e documental, em consonância com a investigação de doutoramento em curso sobre a temática dos tutores virtuais. Como resultados, a literatura especializada tem destacado a enorme vantagem do uso de <i>chatbots</i> (assistentes virtuais) na EaD no ensino superior online, devido aos enormes benefícios para os intervenientes diretos do processo de ensino-aprendizagem, ou seja: a instituição, os tutores e os alunos.
Conteúdos	
Outro título	
Títulos relacionados	
Séries	
Editor	Universidade de Passo Fundo (UPF)
Data de criação	
Edição	
Formato	
Frequência	
Fonte	Alma/SFX Local Collectionundefined DOAJ Directory of Open Access Journals - Not for CDI Discovery
Encadernar com	
Permalink	https://unesp.primo.exlibrisgroup.com/permalink/55UNESP_INST/c1988b/cdi_doaj_primary_oai_doaj_org_article_ef8236acb2474c25bdf32914b87ec99f

Fonte: elaborado pelo autor

Outro exemplo citado, é o Repositório BENANCIB que disponibiliza um conjunto de trabalhos e palestras dos Encontros Nacionais de Pesquisa e Pós-Graduação em Ciência da Informação (ENANCIB), desde sua primeira edição, em 1994. Foi desenvolvido no âmbito do Projeto de Pesquisa "Questões em Rede" criado pelo Grupo de Pesquisa “Informação, Discurso e Memória”, da Universidade Federal Fluminense (UFF), onde está hospedado, e tem a parceria da ANCIB. Seu endereço de acesso é por meio da página <http://repositorios.questoesemrede.uff.br/repositorios>, mostrado na Figura 12.

Figura 12 – Página de Busca do BENANCIB

Questões em Rede - Coleções

Este é o Repositório do Projeto de Pesquisa "Questões em Rede". Nele hoje temos disponíveis para busca os trabalhos e palestras apresentados nos Encontros Nacionais de Pesquisa e Pós-Graduação em Ciência da Informação - ENANCIBs, desde sua primeira edição em 1994 na Coleção BENANCIB.

Criado pelo Grupo de Pesquisa "Informação, Discurso e Memória", da Universidade Federal Fluminense, cadastrado no CNPq, em parceria com a ANCIB, é viabilizado por financiamento da Fundação Carlos Chagas de Amparo à Pesquisa do Rio de Janeiro (FAPERJ), Bolsas do CNPq e pelo apoio técnico da Superintendência de Tecnologia da Informação – STI/ UFF, onde também está hospedado.

Encontra-se em fase de testes, e sua disponibilização para a comunidade de pesquisadores, prevista para junho, foi antecipada para comemoração dos 25 anos da Associação Nacional de Pesquisa e Pós Graduação em Ciência da Informação (ANCIB).

Por ser uma versão beta, solicita-se envio de sugestões, relatos de problemas etc. para o e-mail: benancib@vm.uff.br.

Communities in DSpace

Select a community to browse its collections.

- Repositórios

Search DSpace

Go

Advanced Search

Browse

All of DSpace
Communities & Collections
By Issue Date
Authors
Titles
Subjects
Subject

My Account

Login
Register

Fonte: elaborado pelo autor

Quando um usuário faz a pesquisa por algum documento que foi publicado no evento e disponibilizado por meio desta plataforma, na página de resposta são disponibilizadas algumas informações sobre a obra, como título, URI (*Uniform Resource Identifier*), Data, Resumo, e uma forma de se fazer o *download* do arquivo, mas o sistema ainda disponibiliza um *link* para “Mostrar registro completo” onde é possível obter muitos metadados a respeito dessa obra e descritos utilizando o formato *Dublin Core*, como o exemplo e pode ser verificado pelo Quadro 5.

Quadro 5 – Metadados extraídos da página do BENANCIB

BENANCIB		
dc.contributor.author	BEDIN, Jéssica	
dc.contributor.author	CHAGAS, Magda Teixeira	
dc.date.accessioned	2017-04-25T19:41:30Z	
dc.date.available	2017-04-25T19:41:30Z	
dc.date.issued	42850	
dc.identifier.uri	http://hdl.handle.net/123456789/3251	
dc.description.abstract	This paper, seeks to answer the following issues: How librarians of private schools in Florianópolis/SC may contribute to the training of research students in high school? To answer this question aims to investigate the role of the school librarian in the formation of student researchers. As a methodology it is an exploratory and descriptive research with a qualitative approach and can be classified as a literatu research, document and case study. As a result, we expect to identify actions of librarians in the development of skills through school research and give visibility to the role of the librarian and the school library.	pt_BR
dc.language.iso	pt_BR	pt_BR
dc.subject	Bibliotecário Escolar	pt_BR
dc.subject	Biblioteca Escolar	pt_BR
dc.subject	Pesquisa Escolar	pt_BR
dc.title	A atuação do bibliotecário escolar na formação de estudantes pesquisadores no ensino médio	pt_BR

dc.title.alternative	The acting the school librarian the students researchers training in high school	pt_BR
dc.type	Trabalho Poster	pt_BR
dc.ano.evento	2016	pt_BR
dc.cidade.evento	Salvador	pt_BR
dc.edicao.evento	17	pt_BR
dc.numero.gt	3	pt_BR
dc.titulo.gt	Mediação, Circulação e Apropriação da Informação	pt_BR
dc.keywords	School Librarian	pt_BR
dc.keywords	School Library	pt_BR
dc.keywords	School Research	pt_BR
dc.resumo	Neste trabalho, busca-se resposta para a seguinte problemática: Como os bibliotecários das escolas particulares de Florianópolis/SC podem colaborar para a formação de estudantes pesquisadores, no ensino médio? Para responder a essa questão, objetiva-se investigar a atuação do bibliotecário escolar na formação de estudantes pesquisadores. Quanto à metodologia, trata-se de uma pesquisa exploratória e descritiva, com abordagem qualitativa e pode ser classificada como uma pesquisa bibliográfica, documental e estudo de caso. Como resultados, espera-se identificar ações dos bibliotecários no desenvolvimento de habilidades por meio da pesquisa escolar e dar visibilidade para a atuação do bibliotecário e da biblioteca escolar.	pt_BR
dc.referencias	CAMPELLO, B. S. Biblioteca escolar: conhecimentos que sustentam a prática. Belo Horizonte: Autêntica, 2012. CAMPELLO, B. S. et al. Biblioteca escolar como espaço de produção do conhecimento: parâmetros para bibliotecas escolares. Belo Horizonte: Autêntica, 2010. DURBAN ROCA, G. Biblioteca escolar hoje: recurso estratégico para a escola. Porto Alegre: Penso, 2012. FEDERAÇÃO INTERNACIONAL DE ASSOCIAÇÕES DE BIBLIOTECAS E INSTITUIÇÕES. Manifesto para as Bibliotecas Escolares. 2000. Disponível em: <http://www.ifla.org/files/assets/school-libraries-resource-centers/publications/school-library-guidelines/school-library-guidelines-pt.pdf>. Acesso em: 15 jun. 2016. GIL, A. C. Como elaborar projetos de pesquisa. São Paulo: Editora Atlas, 2010. GIL, A. C. Métodos e técnicas de pesquisa social. São Paulo: Atlas, 2008. LANZI, L. A. C.; VIDOTTI, S. A. B. G.; FERNEDA, E. A biblioteca escolar e a geração nativos digitais: construindo novas relações. São Paulo: Cultura Acadêmica, 2013. Disponível em: <http://goo.gl/kew2H4>. Acesso em: 8 jun. 2016. MINAYO, M. C. S. O desafio da pesquisa social. In: MINAYO, M. C. de S.; DESLANDES, S. F.; GOMES, R. Pesquisa social: teoria, método e criatividade. Petrópolis: Vozes, 2009. p. 9-30. MORO, E. L. da S.; ESTABEL, L. B. A pesquisa escolar propiciando a integração dos atores – alunos, educadores e bibliotecários – irradiando o benefício coletivo e a cidadania em um ambiente de aprendizagem mediado por computador. Revista Renote: Novas Tecnologias na Educação, Porto Alegre, v.2, n.1, p.1-10, 2004. Disponível em: <http://seer.ufrgs.br/index.php/renote/article/view/13662>. Acesso em: 5 abr. 2016. PIERUCCINI, I. A busca do conhecimento na escola: a pesquisa escolar e a construção do conhecimento. Salto para o Futuro. Rio de Janeiro, 2007. RODRIGUES, Â. B. L. A biblioteca escolar como diferencial na compra dos serviços educacionais. In: Seminário Biblioteca Escolar: espaço de ação pedagógica, 3., 2004, Belo Horizonte. Anais eletrônicos... Belo Horizonte: GEBE, 2005. p. 31-49. Disponível em: <http://gebe.eci.ufmg.br/downloads/322.pdf>. Acesso em: 1 maio 2016. SILVA, A. S. R.; NEVES, D. A. B.; GOMES, M. Y. F. S. F. Avaliação da biblioteca escolar para o desenvolvimento de competências informacionais: a experiência da biblioteca do Instituto Federal da Bahia - campus Camaçari. Biblioteca Escolar em Revista, Ribeirão Preto, v.2, n.1, p. 20-40,	pt_BR

	2013. Disponível em: <http://www.revistas.usp.br/berev/article/view/106584>. Acesso em: 28 jun. 2016. SILVA, J. D. O. da; CUNHA, J. de A. O papel educativo da biblioteca escolar no contexto do Plano Nacional de Educação. Encontros Bibli, Florianópolis, v.21, n.46, p. 45-58, 2016. Disponível em: <https://periodicos.ufsc.br/index.php/eb/article/view/38695>. Acesso em: 17 jun. 2016.	
dc.como.citar	BEDIN, Jéssica; CHAGAS, Magda Teixeira. A atuação do bibliotecário escolar na formação de estudantes pesquisadores no ensino médio. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 17., 2016, Salvador. Anais... Salvador: UFBA, 2016.	pt_BR

Fonte: elaborado pelo autor

A Biblioteca Digital Brasileira de Teses e Dissertações promove a integração e a disseminação por meio de seu portal de buscas, de trabalhos provenientes dos textos completos das teses e dissertações defendidas nas instituições brasileiras de ensino e pesquisa. Com a missão de contribuir com o aumento da visibilidade da produção científica nacional e a difusão de informações de interesse científico e tecnológico para a sociedade, mantém o portal e permite todo o acesso de forma gratuita. Por outro lado, procura proporcionar transparência a respeito dos investimentos públicos para o progresso da ciência realizado nos programas de pós-graduação. Para ter acesso aos documentos, a entrada é feita por meio de sua página na Internet no endereço <https://bdtd.ibict.br>, mostrado na Figura 13.

Figura 13 – Página de Busca do BDTD

ACESSO E VISIBILIDADE ÀS
TESES E DISSERTAÇÕES
BRASILEIRAS

128 Instituições 517.698 Dissertações 196.580 Teses 714.277 Documentos

Fonte: elaborado pelo autor

Uma das funções dos metadados é padronizar as informações, nesse ponto a BDTD garantiu a utilização de um padrão e assim alcançou a interoperabilidade entre os sistemas de informação das entidades envolvidas. O banco de teses utiliza o Padrão Brasileiro de Metadados para Descrição de Teses e Dissertações (MTD-BR), que foi desenvolvido analisando outros

padrões internacionais para a descrição desse tipo de documento, como o *Electronic Thesis and Dissertations Metadata Standard* (ETD-MS). Hoje é utilizada a versão mais recente vigente, a terceira versão do Padrão, o MTD3-BR, onde foram feitos avanços, acompanhando as mudanças internacionais, e a inclusão de novos sistemas e formatos.

Ao permitir o usuário pesquisar a base de dados, é entregue uma página Web com a resposta da pesquisa e nela já é encontrado diretamente uma série de informações a respeito da obra, como mostrado no Quadro 6, que é um compilado de um exemplo de todas as informações na ordem em que aparecem na página.

Quadro 6 – Metadados extraídos da página do BDTD

Biblioteca Digital Brasileira de Teses e Dissertações (BDTD)	
Título:	Robô humano: estudo sobre humanização no atendimento com <i>chatbot</i>
Parte do Resumo Inglês:	This study addresses how the influence of the persona helps in the humanization of <i>chatbots</i> and seeks to evaluate/measure its impacts in terms of satisfaction with the service provided to the client, based on perceived usefulness and ease of use,. The theoretical review includes definitions of artif...
Nível de Acesso:	openAccess
Data de Defesa:	2020
Autor/a:	Rebecchi, Adriana de Barros lattes
Orientador/a:	Rocha, Thelma Valéria
Banca:	Strehlau, Suzane, Francisco, Eduardo de Rezende
Tipo Documento:	Dissertação
Idioma:	por
Instituição de Defesa:	Escola Superior de Propaganda e Marketing
Programa:	Programa de Mestrado Profissional em Comportamento do Consumidor
Departamento:	ESPM:Pós-Graduação Stricto Sensu
Assuntos em Português:	comportamento do consumidor; <i>chatbot</i> ; inteligência artificial; atendimento a clientes
Assuntos em Inglês:	consumer behavior; <i>chatbots</i> ; artificial intelligence; customer service
Áreas de Conhecimento:	CIENCIAS SOCIAIS APLICADAS > ADMINISTRACAO
Download Texto Completo:	http://tede2.espm.br/handle/tede/506
Descrição	
Citação:	Rebecchi, Adriana de Barros. Robô humano: estudo sobre humanização no atendimento com <i>chatbot</i> . 2020. [58 f.]. Dissertação (Programa de Mestrado Profissional em Comportamento do Consumidor) - Escola Superior de Propaganda e Marketing, [São Paulo] .
Resumo Português:	Este estudo aborda a influência da persona na humanização do <i>chatbot</i> e busca avaliar/mensurar os impactos em termos de satisfação do atendimento prestado ao cliente, a partir da utilidade e facilidade percebidas, analisando seus impactos sob diferentes perspectivas, como idade, escolaridade e tema do atendimento. A revisão teórica inclui definições sobre inteligência artificial, atendimento ao cliente com a ferramenta <i>chatbot</i> e humanização com assistentes virtuais. Como método, optou-se por desenvolver uma análise quantitativa de uma base de atendimentos realizados por <i>chatbot</i> da instituição SEBRAE-SP, com uma survey com 153 clientes atendidos pela mesma ferramenta, para responder às seguintes hipóteses: H1 – A utilidade percebida

	na experiência com o <i>chatbot</i> impacta diretamente o índice de satisfação. H2 – A facilidade percebida de uso na experiência com o <i>chatbot</i> impacta diretamente o índice de satisfação. H3 – A necessidade de interação humana tem um efeito moderador no impacto da utilidade percebida sobre o índice de satisfação. H4 - A necessidade de interação humana tem um efeito moderador no impacto da facilidade percebida de uso sobre o índice de satisfação. Como resultado, foi comprovado que a utilidade e facilidade de uso percebidas impactam diretamente o índice de satisfação, considerando que a humanização no <i>chatbot</i> durante o atendimento é muito importante para aumentar esse índice. Este estudo traz insights para as empresas que buscam aprimorar seu atendimento com esse tipo de tecnologia, trazendo melhor experiência ao cliente, respeitando a estratégia da empresa.
Resumo inglês:	This study addresses how the influence of the persona helps in the humanization of <i>chatbots</i> and seeks to evaluate/measure its impacts in terms of satisfaction with the service provided to the client, based on perceived usefulness and ease of use, The theoretical review includes definitions of artificial intelligence, customer service with the <i>chatbot</i> tool, and humanization with virtual assistants. As a method, we chose to develop a quantitative analysis of a database of calls performed by <i>chatbot</i> from the institution SEBRAE-SP, through a survey with 153 clients who used this tool. We tested the following hypotheses: H1 - Perceived utility in the experience with a <i>chatbot</i> directly impacts the satisfaction index. H2 – Perceived ease of use in the experience with a <i>chatbot</i> directly impacts the satisfaction index. H3 – The need for human interaction has a moderating effect on the impact of perceived utility on the satisfaction index. H4 - The need for human interaction has a moderating effect on the impact of perceived ease of use on the satisfaction index. The results proved that perceived usefulness and ease of use directly impact the satisfaction index, considering that humanization of the <i>chatbot</i> is very important to increase this index. This study brings insights to companies seeking to improve their service with this type of technology, supplying better customer experience, and respecting the company's strategy.

Fonte: elaborado pelo autor

Estes foram alguns exemplos dos metadados aplicados na Web com a proposta principal de interoperabilidade e disponibilidade, pois ao fornecer as informações a respeito das obras com clareza, permite aos usuários navegarem pelas informações na procura do que pode satisfazer sua necessidade, e ao mesmo tempo, por utilizar etiquetas padronizadas, como no caso do *Dublin Core*, permite que as máquinas possam também navegar pelo conteúdo.

Além disso, fornece condições para que terceiros possam construir ferramentas que podem se beneficiar dessas informações, fato este que vai ao encontro desta pesquisa, a possibilidade de uma tecnologia nova se beneficiar de outra tecnologia já amplamente divulgada e usada com o propósito de ajudar os usuários a encontrar o que procuram, por novas formas de interface, em dados já armazenados e acessados por outras plataformas. Essa é a ideia da interoperabilidade, o reuso dos dados e novas condições de acesso e recuperação.

5.

MODELO PROPOSTO

Este capítulo descreve um modelo baseado na tese que orienta esta pesquisa de que um assistente digital ou um *chatbot*, pode ser utilizado como um sistema de recuperação de informação se este tem acesso aos metadados de um repositório durante o processo de construção dos diálogos para interação com o usuário. Ao criar um repositório de documentos e disponibilizar acesso aos metadados destes, permite-se que ferramentas externas utilizem estes dados como um vocabulário controlado deste domínio. Desta forma, pode-se utilizar novos termos juntamente com os do usuário quando o sistema constrói as expressões de busca, por meio do uso de um sistema baseado em linguagem natural e de uma plataforma de rede social, ou mensageiro instantâneo, o que permite recuperar documentos mais relevantes em um repositório. Os documentos podem ser acessados por meio de uma página Web e de um buscador próprio, mas a partir do momento que é introduzida uma nova interface, na forma de um *chatbot*, é possível melhorar a acessibilidade e permite que um número maior de usuários possa utilizar este sistema, além de abranger outros tipos de pessoas, como os portadores de deficiência visual, que podem se beneficiar deste novo tipo de interação.

A proposta utiliza-se da área da Computação Cognitiva, disponibilizadas pelas grandes empresas de tecnologia na forma de *frameworks* de desenvolvimento de agentes conversacionais, aplicada à Bancos de Dados de repositórios, existentes e disponibilizados, com o propósito de colaborar com a área da Recuperação de Informação e a Ciência da Informação. Este modelo foi desenvolvido para que possa ser visualizado de duas formas, primeiramente aplicando o conceito a um grande repositório de documentos disponível na Web, porém de terceiros, para demonstrar que já existem os recursos e são bastante difundidos, e posteriormente sobre um protótipo de repositório construído especificamente para esta pesquisa, nos mesmos moldes, para prova de conceito do modelo. Como exemplo de acesso à

documentos, e seus metadados, existentes e disponibilizados na Web, foi escolhido para análise e construção do modelo o sítio da Biblioteca Digital Brasileira de Teses e Dissertações⁸².

Para o protótipo, e prova de conceito, foi desenvolvido um repositório a partir de dados oriundos de trabalhos completos e resumos expandidos submetidos por pesquisadores, pós-graduados e pós-graduandos *stricto sensu* apresentados em todas as edições do ENANCIB e disponibilizados na Web em diversos sítios, separados por ano do evento e compreendendo de 1994 a 2021, onde todos os dados foram compilados e inseridos em um banco de dados relacional, padrão SQL, originou o protótipo do repositório denominado ENANCIB WEB⁸³. O banco de dados do ENANCIB WEB será utilizado para fornecer os metadados consumidos pelo protótipo do agente conversacional.

Quando um usuário acessa uma plataforma de rede social, aplicativo de mensagens instantâneas, ou faz uma pesquisa por meio de um buscador na Web, muitos destes usuários interagem utilizando frases em linguagem natural e que contêm, além dos termos que deseja pesquisar, uma série de outras informações. Esses sistemas precisam identificar os termos relevantes para a pesquisa, e desprezar outros termos irrelevantes, antes de efetuar a primeira consulta a uma base de dados. Na Figura 14, tem-se um exemplo de interação entre um usuário humano e um sistema de busca de documentos por meio de linguagem natural.

Figura 14 – Exemplo de diálogo entre um humano e um *chatbot*



Fonte: elaborado pelo autor

⁸² Disponível em: <https://bdtd.ibict.br/vufind>. Acesso em: 07 jan. 2022.

⁸³ Disponível em: <http://www.ricdon.com.br/enancib>. Acesso em: 07 jan. 2022.

Este usuário ao interagir com a plataforma por meio de um sistema de diálogo, o faz como se estivesse conversando com outro usuário, inclusive cumprimentando-o, e depois faz um pedido com alguns termos importantes. Os sistemas hoje disponibilizados pela Ciência da Computação, e áreas afins, já fornecem ferramentas bastante eficientes para identificar os termos de um diálogo, com a utilização da inteligência artificial, aprendizado de máquina e processamento de linguagem natural, empresas como a IBM, Microsoft e Google oferecem serviços bastante elaborados para a utilização neste tipo de aplicação. Dessa forma, não é o escopo deste trabalho analisar como é feito o tratamento dos termos e como essas empresas utilizam essas tecnologias, pois a forma como é feito é parecida e utiliza muitos conceitos das áreas supracitadas. Uma característica que pode fazer a diferença é o ato de ter acesso ao banco de dados utilizado pela empresa, seus sistemas gerenciais e/ou repositórios de informações disponibilizadas na Web no momento de interagir com seus usuários. Grande parte dos módulos disponibilizados por empresas fornecedoras de serviços de assistentes digitais e *chatbots* podem ser acessados de forma gratuita e até mesmo usados por empresas, desde que respeitem os limites de acesso e interações, mas todas as empresas pesquisadas (Amazon, Google, Microsoft, IBM), no momento deste trabalho, forneciam módulos pagos para serem utilizados como interface entre o sistema de diálogo e o banco de dados da empresa ou usuário.

Ao analisar o diálogo da Figura 14, pode-se fazer a extração dos termos e obter conjuntos de informações que trazem valores semânticos, ou não, para a pesquisa dos usuários. Exemplos de materiais irrelevantes para a necessidade do usuário são as palavras:

!!!

Eu queria uma

que fale sobre a

com

.

Nesses casos, os termos não expressam nenhuma informação útil a respeito da necessidade de informação do usuário e precisam ser descartados da pesquisa. Inicialmente tem-se um cumprimento “**Olá**” irrelevante para a pesquisa, mas é importante para o diálogo por exprimir um grau de humanidade ao sistema e gerar conforto ao usuário acostumado com diálogos com outros usuários usando a mesma ferramenta, como exemplo se o estão utilizando por meio de uma rede social ou mensageiro instantâneo. Dentre os outros termos, encontra-se

a palavra “**tese**”, se o sistema é pertencente a um repositório que contém este tipo de documento, pode haver um metadado que classifica os trabalhos como tal e assim expressa o tipo de documento da necessidade do usuário, ou seja, ele pesquisa por teses sobre algum assunto. Para determinar o assunto da pesquisa, tem-se outros termos para serem analisados, e até mesmo comparados com os metadados, como “**recuperação de informação**” e “**ontologias**”, estes descrevem o assunto da “**tese**” que o usuário requer. Estes termos podem ter sido retirados de um vocabulário controlado, como um tesouro, no momento da inserção do documento no sistema e o preenchimento dos metadados, o que garante que utilizam algum tipo de padronização e constituem termos de um domínio específico, sendo assim, guardando algum tipo de herança ou hierarquia entre eles.

A partir de uma única frase proferida, diversos termos podem ser utilizados e, assim, é possível começar um diálogo com a apresentação de informações para este usuário. Com o desenvolver do diálogo, todas as dúvidas podem ser sanadas e o sistema capaz de selecionar o menor número de resultados, e que sejam relevantes, pois diferente de um buscador na Web que pode apresentar milhões de resultados e deixar que o usuário decida o que é relevante, apesar de ter sido analisado pelo algoritmo da plataforma e algum tipo de relevância já ter sido utilizada. Um sistema de mensagens instantâneas é utilizado por meio de linguagem natural e é considerado eficiente quando é capaz de extrair o máximo de informações do usuário com o mínimo de interações, de forma ser capaz de apresentar o menor número de resultados, mas que sejam compatíveis com a necessidade informacional e tenham maior relevância.

Para simular uma conversa entre um usuário e um *chatbot* disponibilizado por meio de um portal na Web, foi criada uma simulação contendo uma série de questionamentos e suas respectivas respostas, mostradas no Quadro 7. O diálogo estabelecido por meio desse “*chatbot fictício*”, utiliza dados oriundos das informações disponibilizadas na BDTD, extraídos por meio de pesquisas avançadas no buscador fornecido e por refinamento e alterações nos filtros de acordo com a evolução da conversa, na data de 15 de fevereiro de 2022. Com os documentos, e seus metadados, armazenados na plataforma e as ligações entre os termos podendo ser alcançada por meio destes metadados, ambos disponibilizados, a construção de um diálogo é possível a partir do momento em que se introduz ferramentas para o processamento de termos baseados em linguagem natural, que neste caso, este processo foi efetuado de forma manual pelo autor, além da formulação de frases como resultado dos questionamentos.

A demonstração da usabilidade de uma ferramenta conversacional que acessa dados de um repositório é possível por várias formas, se o desenvolvedor tem acesso aos dados

diretamente do banco de dados, pode formular as questões e recuperar os dados de forma mais direta e rápida, mas também é possível por meio de consultas na própria página utilizando um *software* que possa fazer a interface entre a programação do agente e o código da página do repositório, pois como todos os dados estão identificados em um padrão uniforme de metadados, a construção das pesquisas pode ser aplicada a uma consulta, e em seguida o *software* pode analisar a página apresentada e vasculhando no código as etiquetas dos metadados para extrair as informações para apresentar ao usuário por meio do agente. Assim como é feito pelos sistemas dos buscadores na Web, que conseguem extrair grande parte das informações que precisam das páginas disponibilizadas pelos proprietários, sem a supervisão destes. A conversa a seguir adota este segundo modelo apresentado, com todo o funcionamento da interface de integração sendo feito pelo autor manualmente e não utilizando um *software* para tal.

Quadro 7 – Diálogo entre usuário e *chatbot* fictício da BDTD

Personagem	Enunciado	Ato Interno do sistema
Usuário	Olá, tudo bem???	Identifica um cumprimento (olá) (tudo bem)
Sistema	Olá, estou muito bem, espero que esteja bem. O que você deseja?	Responde ao cumprimento e pergunta a intenção do usuário.
Usuário	Eu queria uma tese que fale sobre a recuperação de informação com ontologias.	Identifica o tipo do documento (tese) e os assuntos (recuperação de informação) e (ontologias). Os outros termos são irrelevantes. Utilizando esses termos importantes faz uma busca e responde ao usuário com os primeiros resultados, ponderando a respeito da quantidade de documentos encontrados e fazendo novos questionamentos com o intuito de reduzir aos documentos mais relevantes.
Sistema	Encontramos 86 trabalhos que condizem com a sua busca. Precisamos de mais informações para encontrar um trabalho mais relevante de acordo com os termos pesquisados.	Depois de efetuar uma busca pelo banco de dados do portal, foram encontrados 86 trabalhos, é uma quantidade muito grande para ser apresentada ao usuário, então são apresentadas algumas opções para tentar encontrar uma maneira de diminuir a quantidade de trabalhos e aumentar a relevância dos resultados.
Usuário	Que tipos de informações você encontrou?	O usuário questiona tipos de informações disponíveis, dentre as informações do portal e disponibilizados por meio de metadados, as opções são apresentadas ao usuário.

Sistema	Para refinar a busca temos os seguintes parâmetros: "título" "resumo em português" "resumo em inglês" "autor" "orientadores" "assunto" "instituição" "tipo do documento" "nível de acesso" "ano da defesa" "urls" "formato" "língua"	Apresenta ao usuário opções que podem ser selecionadas com o propósito de afinar os resultados e aumentar a relevância.
Usuário	Eu quero apenas os trabalhos mais recentes do ano de 2020	Foi identificado o termo (ano) que condiz com um dos parâmetros de refino da busca e um valor (2020), é feita uma nova busca, mantendo os termos anteriores e acrescentando o filtro de 2020 à busca, encontrando apenas 2 documentos, como são poucos, são apresentados ao usuário para que ele possa avaliar a relevância.
Sistema	Encontrei 2 trabalhos de 2020: 1 - Recuperação da informação com abordagem semântica utilizando linguagem natural: a inteligência artificial na ciência da informação 2 - Abordagem baseada em ontologia para reconhecimento de oportunidades de negócios em notícias extraídas da Web [recurso eletrônico] 3 - Desejo Refinar mais a minha busca	O sistema aguarda o usuário escolher entre um dos dois documentos apresentados ou escolher a terceira opção onde ele desfaz a última seleção de ano e começa novamente a filtragem.
Usuário	opção 3, quero outras opções	Identificado uma opção (3) e o sistema desfaz a última filtragem e volta a perguntar ao usuário sobre suas intenções.
Sistema	Entendi, você pode me dizer mais informações a respeito do que deseja? Pode ser algum dos parâmetros que já te mostrei.	O sistema aguarda o usuário formular uma nova pergunta ou inserir novas informações.
Usuário	Eu gostaria que o assunto tivesse a ver com a web semântica.	Identificado um parâmetro (assunto) e um valor (web semântica), em seguida é efetuada uma nova pesquisa com a filtragem ativada com o novo termo pesquisado dentro do campo assunto.
Sistema	Encontramos 14 trabalhos, dentre os 86, que condizem com a sua busca. Deseja ter acesso a esses documentos (sim ou não)? ou deseja fazer uma nova pergunta para encontrar um trabalho mais relevante?	Depois de efetuar uma busca pelo banco de dados do portal, foram encontrados 14 documentos, é uma quantidade razoável a ser apresentada ao usuário, caso ele deseje, mas são apresentadas algumas opções para tentar encontrar uma maneira de diminuir a quantidade de trabalhos e

		aumentar a relevância dos resultados.
Usuário	Quais as universidades destes trabalhos?	A partir desses 14 documentos é feita a sumarização dos resultados e é apresentado ao usuário.
Sistema	Os 14 trabalhos são pertencentes às seguintes instituições: 1 - UNICAMP (4 trabalhos) 2 - ITA (2 trabalhos) 3 - UFMG (2 trabalhos) 4 - UNESP (2 trabalhos) 5 - PUC_MINS (1 trabalho) 6 - UFRGS (1 trabalho) 7 - USP (1 trabalho) 8 - UTFPR (1 trabalho) 9 - Desejo Refinar mais a minha busca	O sistema aguarda o usuário escolher entre uma das opções apresentadas ou escolher a última opção onde ele desfaz a última seleção de instituição e começa novamente a filtragem.
Usuário	Fiquei curioso para ver os trabalhos do ITA, me mostre.	Identificado um valor (ITA), que consta como uma opção (2) e em seguida é efetuada uma nova pesquisa com a filtragem ativada com o novo termo pesquisado dentro do campo instituição.
Sistema	Encontrei 2 trabalhos de ITA: 1 - Modelo e método semântico para organização e recuperação de documentos 2 - Análise da relevância semântica na seleção de atributos para a mineração de dados 3 - Desejo Refinar mais a minha busca	O sistema aguarda o usuário escolher entre um dos dois documentos apresentados ou escolher a terceira opção onde ele desfaz a última seleção e começa novamente a filtragem.
Usuário	Interessante o trabalho 1, me fale quais os assuntos que ele aborda.	Identificado um valor (1), que consta como opção (1) e um valor (assuntos) que é um dos parâmetros de metadados e em seguida é efetuada uma pesquisa para descobrir quais os assuntos registrados naquele documento.
Sistema	Os assuntos do trabalho 1 são: Arquitetura de software Web semântica Ontologias (inteligência artificial) Tratamento da informação Recuperação da informação Documentos eletrônicos Engenharia de software Computação	O sistema aguarda o usuário formular uma nova pergunta ou inserir novas informações.

Usuário	Existem outros trabalhos cujo assunto abordado é a arquitetura de software?	Identificado um valor (assunto), que consta como parâmetros de metadados e um valor (arquitetura de software) que é um termo encontrado dentro do campo assunto e em seguida é efetuada uma pesquisa para descobrir quais outros trabalhos tem arquitetura de software como assunto.
Sistema	Não foram encontrados outros trabalhos cuja arquitetura de software é um assunto relacionado.	O sistema aguarda o usuário formular uma nova pergunta ou inserir novas informações.
Usuário	Me mostre novamente os trabalhos do ITA.	Identificado um valor (ITA), que consta como uma opção pesquisada anteriormente e em seguida é efetuada uma pesquisa com a filtragem ativada com o termo dentro do campo instituição.
Sistema	Encontrei 2 trabalhos de ITA: 1 - Modelo e método semântico para organização e recuperação de documentos 2 - Análise da relevância semântica na seleção de atributos para a mineração de dados 3 - Desejo Refinar mais a minha busca	O sistema aguarda o usuário escolher entre um dos dois documentos apresentados ou escolher a terceira opção onde ele desfaz a última seleção e começa novamente a filtragem.
Usuário	Eu quero ver o trabalho 1	Identificado um valor (1), que consta como opção (1) e em seguida é efetuada uma pesquisa para descobrir as informações registradas naquele documento.
Sistema	Para acessar o trabalho "Modelo e método semântico para organização e recuperação de documentos" utilize a seguinte URL: http://www.bd.bibl.ita.br/tde_busca/arquivo.php?codArquivo=2873	O sistema apresenta a URL ao usuário onde ele pode encontrar todas as informações sobre o documento e aguarda o usuário formular uma nova pergunta ou inserir novas informações.
Usuário	Muito Obrigado	Identifica uma despedida/agradecimento (muito obrigado)
Sistema	Eu que agradeço e estou sempre à disposição.	Responde ao agradecimento e se despede.

Fonte: elaborado pelo autor

Pode-se ver na coluna denominada “Ato Interno do sistema”, o que o sistema de análise de interações provenientes de dados oriundos do repositório interpreta a partir do diálogo com um usuário. É possível observar como podem ser formuladas as questões que precisam ser processadas e respondidas, além disso, é possível determinar o que pode ser utilizado pelo sistema de diálogo sem depender de informações externas e interagir com o usuário, ou quando determinar que maiores informações são necessárias, ele pode formular uma consulta ao

repositório para extrair mais dados para prosseguir com o diálogo. Em todo o diálogo, apenas uma vez foi permitido o acesso ao documento completo, e ainda por meio de sua URL, em todas as outras interações com apenas acesso aos metadados é possível formular respostas e novas perguntas para a interação com dados bastante ricos a respeito do assunto tratado neste momento.

A utilização dos metadados para orientar o diálogo permite que informações como ano, autores, assuntos ou palavras-chave possam ser utilizadas na construção das respostas ao usuário na busca de eliminar o máximo de resultados irrelevantes para a pesquisa atual, onde até mesmo a ligação dos termos, que podem ser oriundos de vocabulários controlados, permite entregar uma qualidade bastante superior na qualidade das buscas com o intuito de direcionar o usuário ao que ele precisa com o mínimo de troca de palavras.

Interações como cumprimento, despedida ou agradecimento podem ser tratados por sistemas de construção de *chatbots* de maneira bastante tranquila, pois utilizando sistemas de aprendizagem de máquina ou inteligência artificial, além de regras definidas para estas situações, não são necessários mais dados para formular as respostas e enviá-las aos usuários, não importando o meio de utilização, voz ou texto. Outra situação muito importante que deve ser considerada nestes momentos, em casos em que mais informações são necessárias, é um sistema não configurado corretamente ou apenas ainda não treinado suficientemente, pode ter dificuldades de acompanhar o contexto da conversa para seguir o fluxo de interações. Atualmente existem sistemas automáticos de atendimento a clientes e usuários e por serem programados para apenas seguirem as regras definidas no sistema, não são capazes de acompanhar o desenrolar da conversa de forma satisfatória e acabam fugindo do assunto, dessa forma descontextualizando e causando desconforto ao usuário, terminando por aumentar o número de reclamações das pessoas, quando estes precisam usar sistemas deste tipo.

Como demonstrado anteriormente, o diálogo entre esse usuário e o sistema vai se desenrolando e o sistema vai apresentando opções na forma de linguagem natural, tentando imitar uma conversa com outro ser humano. Muitos sistemas que utilizam inteligência artificial conseguem fazer diálogos muito complexos e detalhados, mas demandam muita tecnologia e conhecimento dos desenvolvedores para criar e manter um sistema assim. Outrora utilizando o modelo proposto neste trabalho, muita desta “inteligência” pode ser conseguida ao alimentar o sistema com os metadados no processo de interação com o usuário, pois são basicamente consultas a um banco para ter acesso às ligações entre os termos existentes.

Estando hospedados no sistema por meio dos metadados, que em alguns casos, podem ser derivados de vocabulários controlados de domínio ou mesmo termos de uma área específica dos documentos, de qualquer forma, os termos possuem um valor semântico bastante significativo a respeito dos documentos a que se referem.

Na busca do entendimento de como sistemas assim propostos podem auxiliar os usuários, por meio das mais diversas modalidades de interação, foi necessária a criação de um banco de dados com informações úteis ao projeto, e para isso foi escolhido os trabalhos apresentados no ENANCIB, em todas as edições, totalizando **5048 trabalhos** encontrados nas modalidades de Trabalho Completo e Resumo Expandido, no momento da construção de tal *corpus* para este trabalho, muitas informações importantes foram encontradas, e se vai relatar algumas delas.

A cada ano do evento é realizado em um local diferente e com temática distinta, conforme pode ser observado no Quadro 8 a seguir:

Quadro 8 – Informações das Edições do ENANCIB

ANO	EDIÇÃO	DATA	INSTITUIÇÃO / CIDADE	TEMA
1994	I	8 a 10 de abril de 1994	UFMG / Belo Horizonte - MG	
1995	II	22 a 24 de novembro de 1995	PUC-Campinas / Valinhos - SP	
1997	III	10 a 12 de setembro de 1997	IBICT/UFRJ / Rio de Janeiro - RJ	
2000	IV	6 a 10 de novembro de 2000	UnB / Brasília - DF	Conhecimento para o Século XXI: a pesquisa na construção da Sociedade da Informação
2003	V	10 a 14 de novembro de 2003	UFMG / Belo Horizonte - MG	Informação, conhecimento e transdisciplinaridade
2005	VI	28 a 30 de novembro de 2005	UFSC / Florianópolis - SC	A política científica e os desafios da sociedade da informação
2006	VII	19 a 22 de novembro de 2006	Unesp / Marília - SP	A dimensão epistemológica da Ciência da Informação e suas interfaces técnicas, políticas e institucionais nos processos de produção, acesso e disseminação da informação
2007	VIII	28 a 31 de outubro de 2007	UFBA / Salvador - BA	Promovendo a inserção internacional da pesquisa brasileira em Ciência da Informação
2008	IX	28 de setembro a 1 de outubro de 2008	USP / São Paulo - SP	Diversidade cultural e políticas de informação
2009	X	25 a 28 de outubro de 2009	UFPB / João Pessoa - PB	A responsabilidade social da Ciência da Informação
2010	XI	25 a 28 de outubro de 2010	IBICT/UFRJ; Fiocruz; UNIRIO / Rio de Janeiro - RJ	Inovação e inclusão social: questões contemporâneas da informação
2011	XII	23 a 26 de outubro de 2011	UnB / Brasília - DF	Políticas de Informação para a Sociedade

2012	XIII	28 a 31 de outubro de 2012	Fiocruz / Rio de Janeiro - RJ	A sociedade em rede para a inovação e o desenvolvimento humano
2013	XIV	29 de outubro a 1 de novembro de 2013	UFSC / Florianópolis - SC	Informação e interação: ampliando perspectivas para o desenvolvimento humano
2014	XV	27 a 31 de outubro de 2014	UFMG / Belo Horizonte - MG	Além das 'nuvens': expandindo as fronteiras da Ciência da Informação
2015	XVI	26 a 30 de outubro de 2015	UFPB / João Pessoa - PB	Informação, Memória e Patrimônio: do documento às redes
2016	XVII	20 a 25 de novembro de 2016	UFBA / Salvador - BA	Descobrimientos da Ciência da Informação: desafios da Multi, Inter e Transdisciplinaridade (MIT)
2017	XVIII	23 a 27 de outubro de 2017	UNESP / Marília - SP	Informação, Sociedade, Complexidade
2018	XIX	22 a 26 de outubro de 2018	UEL / Londrina - PR	O Sujeito Informacional e as Perspectivas Atuais em Ciência da Informação
2019	XX	21 a 25 de outubro de 2019	UFSC / Florianópolis - SC	A Ciência da Informação e a era dos dados
2021	XXI	25 a 29 de outubro de 2021	UFRJ / Rio de Janeiro - RJ	50 anos de Ciência da Informação no Brasil: saberes, diversidade e transformação social
2022	XXII	07 a 11 de novembro de 2022	UFRGS / Porto Alegre - RS	

Fonte: elaborado pelo autor

Além disso, a cada evento é criado um sítio na Web com as informações, as formas de participar e ter acesso aos anais do evento, e com o tempo muitos destes repositórios passaram por mudanças e até mesmo ficaram indisponíveis, além de não ser utilizado um padrão na forma de como as informações são disponibilizadas, isso tudo gera dificuldades em encontrar os dados completos sobre todas as apresentações e até mesmo os documentos publicados. Verificou-se que poucos trabalhos disponibilizados não foram encontrados no formato de PDF, pois não foram disponibilizados por meio de sua URL, seja por sua indisponibilidade ou por serem muito antigos (principalmente dos primeiros eventos), mas ressalta-se que todos os documentos encontrados foram descarregados para serem utilizados e disponibilizados posteriormente no sistema. Outros problemas encontrados foram a utilização de nomes de autores não padronizados pelo fato de não se utilizar nenhum tipo de controle sobre isso, apenas a responsabilidade do usuário em digitar os nomes dos autores, exemplo disso, dentro de um mesmo evento um autor tem seu nome descrito de forma completa e em outro trabalho escrito com sobrenomes abreviados ou escritos incorretamente. Foram encontrados muitos erros de digitação em diversos campos dos dados, fato que pode influenciar diretamente na possibilidade destes trabalhos serem encontrados. Observou-se também que o padrão de publicação de no máximo 5 autores, que consta nos editais de publicação, não foi respeitado em algumas edições do evento, pois foram identificados diversos trabalhos com um número superior de autores, sendo que um caso contém 19 autores cadastrados. É recomendado que sejam utilizados termos

do Tesauro da Ciência da Informação⁸⁴ para a criação das palavras-chaves, mas isto frequentemente não é obedecido, além de não haver um padrão na criação destes termos ainda ocorreram erros de escrita e termos estranhos, como nomes próprios e datas.

Com o intuito de provar o conceito por meio de um protótipo, foi desenvolvido um portal de consulta aos trabalhos apresentados no ENANCIB, pois os dados oriundos da Web trazem muitas informações sobre estas apresentações e foram catalogadas em um banco de dados MySQL⁸⁵ hospedado na própria Web. De princípio foram cadastrados no banco os seguintes dados a respeito de cada documento:

- Ano
- GT_Código
- GT_Nome
- Título
- Palavras-chave (1 a 5)
- Autores (1 a 19)
- Resumo
- *Link para download* do documento em PDF

Estas informações, ou metadados, foram extraídas de forma manual das páginas disponíveis de cada evento, ou instituições parceiras, e cadastrados no banco de dados, durante os anos de 2018 e 2021. Foram selecionados estes campos devido a algumas características desse *corpus*. O ano é importante para entender a qual edição do evento o trabalho pertence e isso deve ser levado bastante em consideração nas pesquisas. Os campos GT_Codigo e GT_Nome são importantes, pois com o decorrer dos anos alguns grupos de trabalhos foram criados ou até mesmo tiveram os seus nomes alterados e dessa forma para garantir a integridade e a coerência dos trabalhos com os respectivos grupos de trabalho. Essas informações foram mantidas e estão relacionadas com os trabalhos de tal forma que não possam ser alteradas caso ocorram novas mudanças nos grupos de trabalho. O título e as palavras-chave expressam muito a respeito do que trata o documento e recebem uma importância elevada no momento da construção de uma busca para recuperar as informações ao usuário. Foi definido que haveria cinco campos para as palavras-chaves e que a ordem destas palavras é definida pelo autor e

⁸⁴ Disponível em: <https://www.gov.br/ibict/pt-br/central-de-conteudos/publicacoes/tesauro-brasileiro-de-ciencia-da-informacao>. Acesso em: 17 fev. 2022.

⁸⁵ Disponível em: <https://www.mysql.com/>. Acesso em: 17 fev. 2022.

deve ser levado em consideração. A respeito do campo que contém os nomes dos autores é importante ressaltar que em algumas edições não foi respeitada a quantidade máxima de 5 autores por trabalho, e para isso foi necessário que este sistema acompanhasse esta “anomalia” de forma a garantir a importância na ordem dos autores gerando 19 colunas de autores no banco de dados. O resumo é disponibilizado ao usuário no momento da sua consulta, pois permite a este verificar se os dados que o documento contém podem ajudá-lo ou não na sua necessidade de informação, antes mesmo de abrir o documento em si, fato este que se dá por meio do *link* para *download* utilizando a URL fornecida, que neste momento ainda aponta para o local original do arquivo sob a guarda da entidade responsável por aquela edição do evento, mas com o intuito de não depender dessa situação no futuro, todos os documentos disponíveis foram descarregados em formato PDF e que, em um trabalho futuro, serão adicionados ao sistema e os usuários poderão ter acesso aos documentos independente se o repositório original da edição estiver disponível ou não, e que pode-se dizer que esta será uma segunda fase deste trabalho.

Na Figura 15, tem-se a página de entrada do sítio ENANCIB WEB que foi disponibilizado por meio do endereço <http://www.ricdon.com.br/enancib> para acessar dados de trabalhos apresentados até a edição de 2021.

Figura 15 – Página de pesquisa do ENANCIB WEB



ENANCIB WEB

Encontro Nacional de Pesquisa em Ciência da Informação

Pesquisa de trabalhos em dados oriundos dos Trabalhos Completos e Resumos Expandidos submetidos por pesquisadores, pós-graduados e pós-graduandos stricto sensu apresentados em todas as edições do evento e disponibilizados na internet

Ano, Título, Resumo, Autor ou palavras chave

Pesquisas realizadas: 139 - © 2020-2022 - Ricardo César de Carvalho - v0.3.1 - [Início](#) - [Contato](#) - [Sobre](#)

Fonte: elaborado pelo autor

Por meio desta página pode ser introduzido qualquer termo, ou conjunto de termos, para serem utilizados na busca de trabalhos apresentados, até o ano de 2021 neste momento, e depois

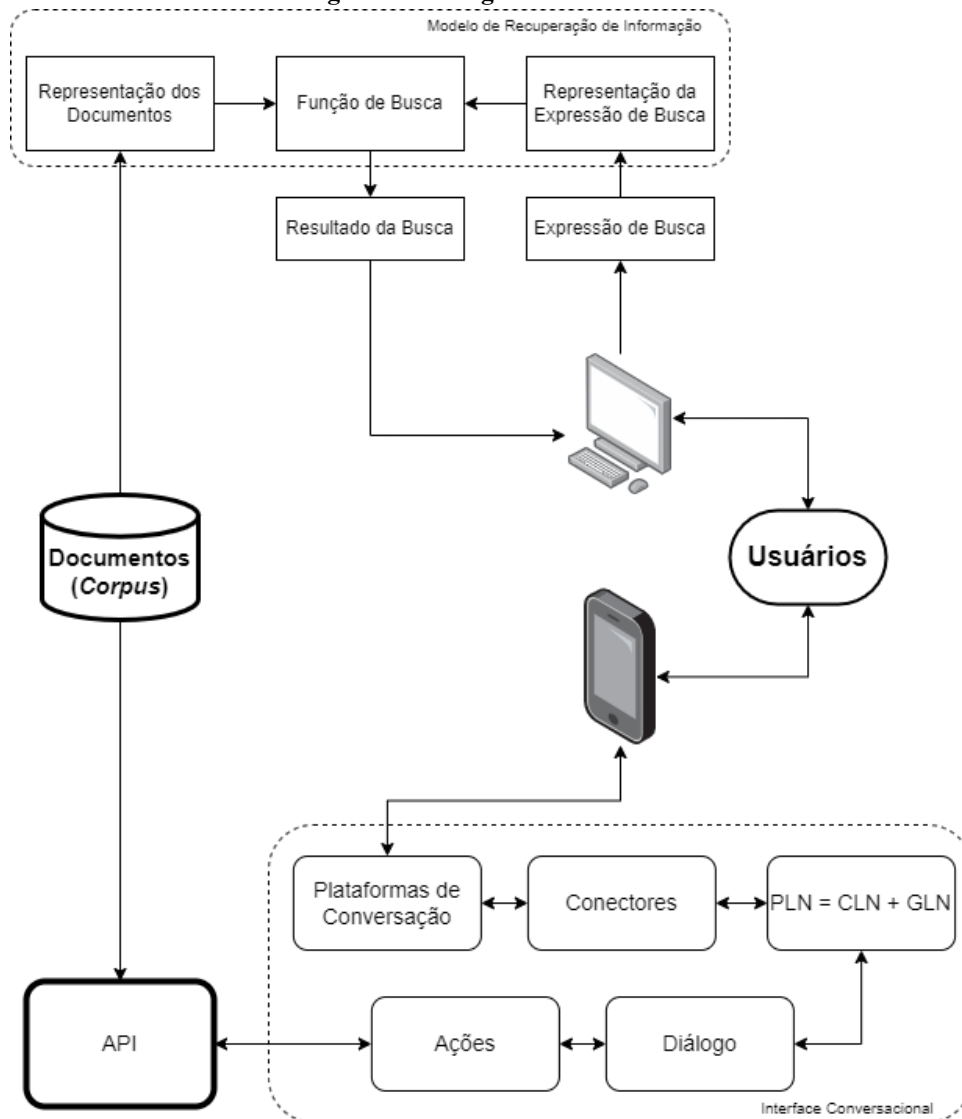
de apresentados os resultados, o usuário pode acessar mais informações sobre cada trabalho e ter acesso a uma URL com a qual poderá fazer o *download* do arquivo em formato PDF, lembrando que estes arquivos são armazenados e publicados por muitas instituições e às vezes as páginas não estão disponíveis, quando é feita uma migração ou por qualquer outro problema.

5.1 Modelo de Interface Conversacional

A partir dos conceitos analisados e como proposta de implementação de um trabalho aplicado na área da Ciência da Informação, um modelo começou a ser desenvolvido e pode ser visualizado na Figura 16.

Partindo do modelo de recuperação de informação de Ferneda (2012), foi proposto um modelo em que uma interface adicional, na forma de *chatbot*, pudesse ser adicionada a um sistema existente, de forma a complementá-lo permitindo uma nova forma de acesso por meio de uma rede social, por exemplo. Mas não necessitando de plataformas complexas, ou caras de inteligência artificial para introduzir essa inteligência de domínio, pois isso pode ser conseguido ao ter acesso aos metadados do sistema existente e que são disponibilizados pelo mantenedor.

Figura 16 – Diagrama do Modelo



Fonte: elaborado pelo autor

O usuário pode acessar as informações disponibilizadas pela empresa por meio de sua página na Web utilizando um computador por meio de uma tela com um campo, podendo utilizar quaisquer termos para procurar algum documento que corresponda à sua necessidade. Neste exemplo é utilizada a página principal do sistema desenvolvido denominado e ENANCIB WEB que permite ao usuário procurar trabalhos publicados no evento de mesmo nome, conforme mostrado na Figura 17. O usuário pesquisa sobre “ciência de dados” e encontra dois trabalhos publicados em 2021 no GT 8, e pode ter acesso a mais informações antes de abrir o documento PDF. Ao abrir o documento tem acesso ao trabalho publicado por meio do *template* do evento e acesso a todas as informações, além dos metadados.

Figura 17 – Utilização do buscador do sítio ENANCIB WEB

The image shows the ENANCIB WEB search interface. At the top, there is a logo of a scientist and the text "ENANCIB WEB". Below it, a description of the site's purpose is provided. A search bar with the text "Ciência de dados" and a "Buscar" button is shown. Below the search bar, a table displays search results. The table has columns for "Ano", "Titulo", and "Nome GT". Two results are shown for the year 2021. A blue arrow points from the "Ver" link of the second result to a detailed view of the event. The detailed view shows the title, authors, keywords, and a link to access the material. A blue arrow points from the "Clique para acessar o material" link to a PDF document titled "XXI ENANCIB".

ENANCIB WEB
Encontro Nacional de Pesquisa em Ciência da Informação

Pesquisa de trabalhos em dados oriundos dos Trabalhos Completos e Resumos Expandidos submetidos por pesquisadores, pós-graduados e pós-graduandos stricto sensu apresentados em todas as edições do evento e disponibilizados na internet

Ciência de dados **Buscar**

Registros encontrados: 2

Ano	Titulo	Nome GT
2021	Análise automatizada do autoarquivamento na Ciência da Informação: um método reprodutível	8 - Informação e Tecnologia
2021	INTELIGÊNCIA ARTIFICIAL E CIÊNCIA DE DADOS EM CRIS INSTITUCIONAL: MODELO CONCEITUAL	8 - Informação e Tecnologia

Pesquisas realizadas: 139 - © 2020-2022 - Ricardo César de Carvalho - v0.3.1 - [Início](#) - [Contato](#) - [Sobre](#)

Detalhes do Evento

Ano: 2021 GT: 8 - Informação e Tecnologia

Titulo
INTELIGÊNCIA ARTIFICIAL E CIÊNCIA DE DADOS EM CRIS INSTITUCIONAL: MODELO CONCEITUAL

Resumo
"A crescente disponibilização de dados e informações relacionadas à ecologia de pesquisa, em múltiplos sistemas de informação, culmina em crescente complexidade na atividade de gestão de pesquisa. Como resposta a esta complexidade, são desenvolvidos Sistemas de Informação de Pesquisa Corrente, do inglês Current Research Information System (CRIS) que objetivam o gerenciamento de metadados contextuais das atividades de pesquisa relacionadas a determinada instituição, quer seja de pesquisa ou de fomento. Neste estudo, um Modelo Conceitual de CRIS institucional é revisitado, visando o aprimoramento do processo, com o uso de Inteligência Artificial e Ciência de Dados. Como procedimento metodológico, utiliza a revisão bibliográfica para o embasamento teórico-conceitual a fim de contextualizar a Inteligência Artificial e a Ciência de Dados, incorporadas ao estudo. A partir disso, criou-se o modelo revisitado, inserindo uma camada de dados que trata dos aspectos da Ciência de Dados, bem como técnicas e métodos de Inteligência Artificial em todos os processos do CRIS, em especial, inseriram-se Processamento de Linguagem Natural, Visão Computacional, Mineração de Texto e Aprendizagem de Máquina. Conclui-se, assim, que a adaptação apresentada do modelo se mostra como um amadurecimento na própria compreensão que se tem do CRIS, com a inserção de elementos que tornam tal modelo mais atualizado. Assim, os estudos e a criação de modelos e de aplicações de CRIS permitem uma evolução na gestão institucional."

Autores
Caio Saraiva Coneglian; Emanuelle Torino; Silvana Aparecida Bonsetti Gregorio Vidotti

Palavras chave
CRIS; Inteligência Artificial; Ciência de Dados; Gestão de Pesquisa

Link
[Clique para acessar o material](#)

XXI ENANCIB
Encontro Nacional de Pesquisa em Ciência da Informação - 2022

874 - Informação e Tecnologia

INTELIGÊNCIA ARTIFICIAL E CIÊNCIA DE DADOS EM CRIS INSTITUCIONAL: MODELO CONCEITUAL

ARTIGO AUTORIZADO PARA USO EXCLUSIVO EM CRIS INSTITUCIONAL: MODELO CONCEITUAL

Caio Saraiva Coneglian - Centro Universitário Fundação de Marília (UNIFAMAR)
Emanuelle Torino - Universidade Tecnológica Federal do Paraná (UTFPR), Universidade
Silvana Aparecida Bonsetti Gregorio Vidotti - Universidade Estadual Paulista (UNESP)

Resumo
Resumo: A crescente disponibilização de dados e informações relacionadas à ecologia de pesquisa, em múltiplos sistemas de informação, culmina em crescente complexidade na atividade de gestão de pesquisa. Como resposta a esta complexidade, são desenvolvidos Sistemas de Informação de Pesquisa Corrente, do inglês Current Research Information System (CRIS) que objetivam o gerenciamento de metadados contextuais das atividades de pesquisa relacionadas a determinada instituição, quer seja de pesquisa ou de fomento. Neste estudo, um Modelo Conceitual de CRIS institucional é revisitado, visando o aprimoramento do processo, com o uso de Inteligência Artificial e Ciência de Dados. Como procedimento metodológico, utiliza a revisão bibliográfica para o embasamento teórico-conceitual a fim de contextualizar a Inteligência Artificial e a Ciência de Dados, incorporadas ao estudo. A partir disso, criou-se o modelo revisitado, inserindo uma camada de dados que trata dos aspectos da Ciência de Dados, bem como técnicas e métodos de Inteligência Artificial em todos os processos do CRIS, em especial, inseriram-se Processamento de Linguagem Natural, Visão Computacional, Mineração de Texto e Aprendizagem de Máquina. Conclui-se, assim, que a adaptação apresentada do modelo se mostra como um amadurecimento na própria compreensão que se tem do CRIS, com a inserção de elementos que tornam tal modelo mais atualizado. Assim, os estudos e a criação de modelos e de aplicações de CRIS permitem uma evolução na gestão institucional."

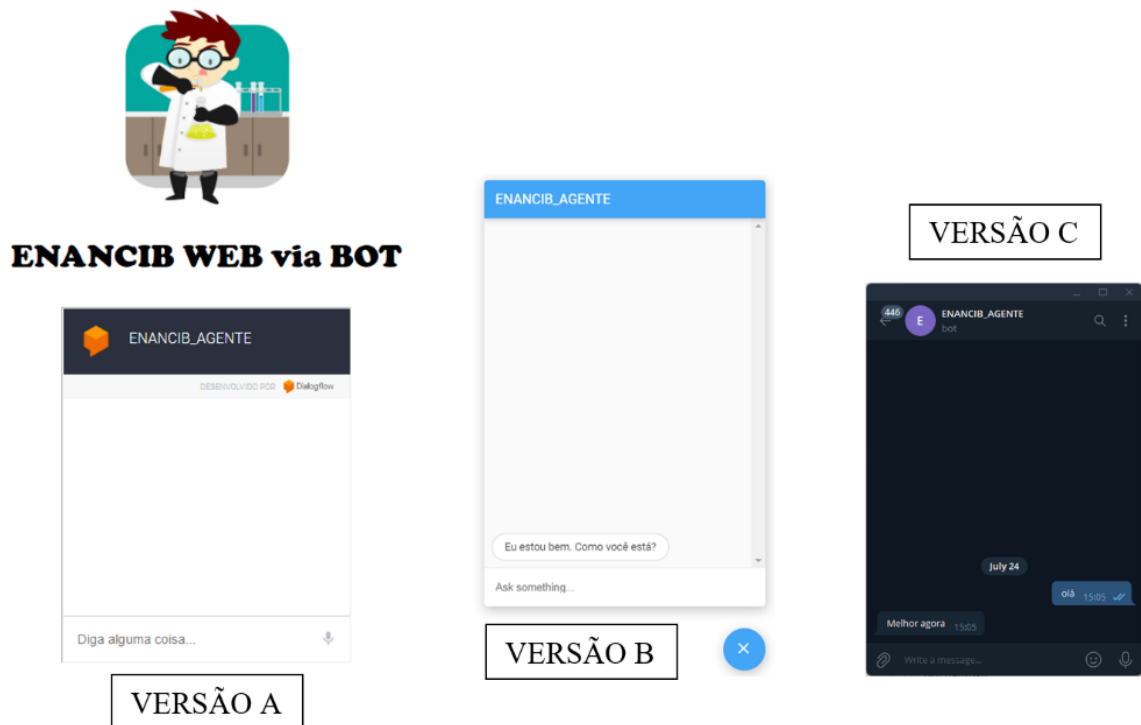
Palavras-chave
CRIS; Inteligência Artificial; Ciência de Dados; Gestão de Pesquisa

Abstract
Abstract: The increasing availability of data and information related to research ecology, in multiple information systems, culminates in increasing complexity in the activity of research management. As a response to this complexity, Current Research Information Systems (CRIS) are developed, which aim to manage the contextual metadata of research activities, whether for research or for funding. In this study, a Conceptual Model of CRIS institutional is revisited, aiming to improve the process, with the use of Artificial Intelligence and Data Science. As a methodological procedure, it uses bibliographic review to contextualize Artificial Intelligence and Data Science, incorporated into the study. From this, the revised model was created, inserting a layer of data that deals with the aspects of Data Science, as well as techniques and methods of Artificial Intelligence in all the processes of the CRIS, in particular, Natural Language Processing, Computer Vision, Text Mining and Machine Learning were inserted. It is concluded, therefore, that the adaptation presented of the model shows as a maturation in the own understanding that one has of the CRIS, with the insertion of elements that make the model more up-to-date. Thus, the studies and the creation of models and applications of CRIS allow for an evolution in institutional management."

Fonte: elaborado pelo autor

É considerado um sistema de recuperação de informação como qualquer outro disponibilizados na Web para os mais diversos fins. Contudo, a partir deste momento é possível utilizar as informações cadastradas em seu banco de dados para alimentar outro sistema com uma interface diferente para o mesmo *corpus*. A Figura 18 nos mostra diferentes interfaces que utilizam o modelo de *chatbot* e que permite acesso à mesma base de dados, porém desta forma utilizando interface com linguagem natural escrita ou falada.

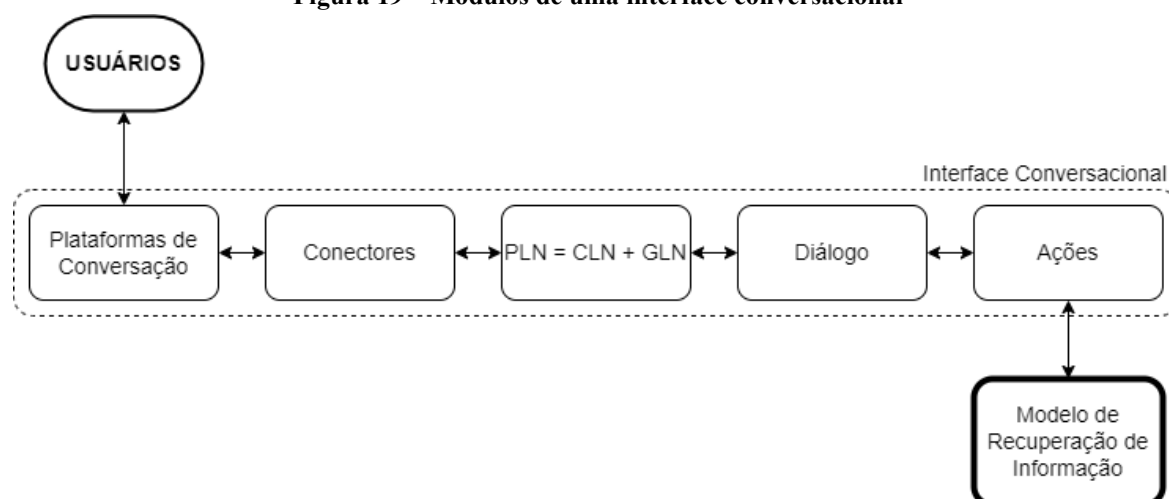
Figura 18 – Interfaces conversacionais do ENANCIB WEB



Fonte: elaborado pelo autor

A versão A é disponibilizada numa página Web própria ou pode ser acessada em qualquer página do repositório na forma de um pequeno ícone disponível em algum canto da página, a versão B é utilizada por meio do aplicativo de mensagens Telegram em um computador pessoal e a versão C é utilizada por meio do aplicativo de mensagens Telegram em um celular Android. Para desempenhar satisfatoriamente a função de integrar um sistema baseado em um banco de dados a uma interface por meio de um *chatbot*, esta interface pode ser dividida em vários módulos, mostrado na Figura 19, e cada módulo tem uma função bem definida, e muitos deles, o desenvolvedor nem precisa de acesso a nível de programador porque são fornecidos na forma de pacotes fechados e já configurados para fazer a sua função. Nesses casos, a maior demanda do desenvolvedor fica a cargo do módulo de AÇÕES, onde é possível fazer a interligação entre esta plataforma e algum outro sistema existente.

Figura 19 – Módulos de uma interface conversacional



Fonte: elaborado pelo autor

Ao dividir o processo em módulos, é possível permitir ao desenvolvedor separar os processos de acordo com a sua interação com os dados. Algumas ferramentas utilizadas na criação de *chatbots* utilizam nomes diferentes para estes processos, mas no final os processos internos de análise e processamento dos dados é muito parecido, no exemplo, utiliza-se a ferramenta do Dialogflow disponibilizada pelo Google, que será descrita com maiores detalhes adiante. No Quadro 9 são apresentados os módulos da interface conversacional e suas respectivas ações desempenhadas.

Quadro 9 – Módulos e Ações da Interface Conversacional

Módulo	Ações
Usuário	Usuário tem uma necessidade de informação e decide fazer sua pesquisa por meio de um <i>chatbot</i> disponível numa página Web ou aplicativo de redes sociais ou mensagens.
Plataformas de Conversação	Um usuário utilizando uma Plataforma de Conversação, como uma rede social ou canal disponibilizado pela empresa para contato, envia uma mensagem que é direcionada até o servidor do <i>chatbot</i> .
Conectores	No servidor, os pacotes de dados são recepcionados por um Conector, que precisa ser específico para cada Plataforma de Conversação. Neste momento os dados provindos da rede social, por exemplo, são traduzidos para que possam ser processados, separando o nome do usuário da mensagem enviada.
PLN = CLN + GLN	Depois que os dados já foram tratados, são enviados para o módulo de PLN para que este possa identificar diversas partes como as perguntas, objetivos, contextos, sentido e outros parâmetros que podem estar inseridos na fala. Atualmente várias soluções utilizam a Inteligência Artificial neste módulo para processar a extração dos termos.

Diálogo	O módulo de Diálogo recebe as informações e os termos encontrados e processa o que é necessário obter para formular uma resposta condizente com a pergunta do usuário, neste momento, ele cria uma mensagem de resposta e verifica se é necessário que algum dado precisa ser recuperado de outra fonte.
Ações	Requerendo ao módulo de Ações, que é responsável por coletar informações ou realizar diversas tarefas, como buscas em bancos de dados, APIs, ferramentas de avaliação ou qualquer fonte de dados atribuída para fornecer informações para nosso <i>chatbot</i> . Neste momento, utiliza os termos encontrados, na etapa anterior, para fazer uma busca no vocabulário (metadados) e encontrar mais termos para em seguida construir a expressão de busca e começar o processo da recuperação de informação.
MRI (Modelo de Recuperação de Informação)	Processo de recuperação de informação em uma base de documentos ou mesmo em um banco de dados. Pode ser utilizado um buscador na Web, mas nesse caso pode ser um serviço de terceiro que fará a interface por meio da expressão de busca e do resultado de busca com o módulo de ações.

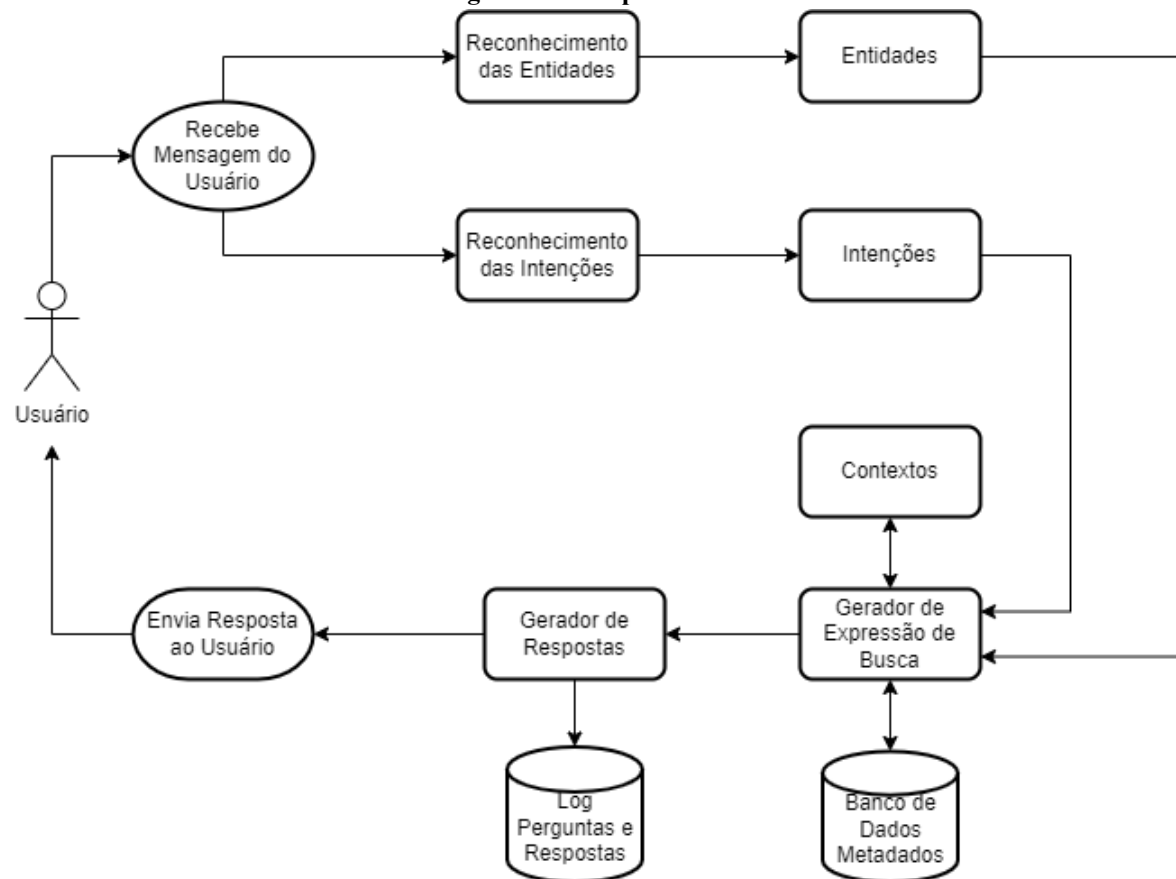
Fonte: elaborado pelo autor

Como descrito no Quadro 9, diversos módulos são disponibilizados pelos fornecedores das plataformas de desenvolvimento em forma de pacotes, nos quais é possível construir *chatbots* de maneira simples, desde que eles não precisem acessar bases de dados de outros fornecedores, pois nesse caso, as ferramentas são pagas e complexas para configuração pelo usuário, além de demandar o trabalho de desenvolvedores para disponibilizar a integração tanto pelo lado do *chatbot*, quanto pelo lado do sistema da empresa. É possível construir agentes com regras definidas de forma simples, onde o usuário ao conversar com este sistema fica navegando entre todas as opções que são apresentadas e pode ter acesso a muitas informações, mas quando é necessário acesso a informações personalizadas ou provindas de bancos de dados, uma nova camada é introduzida na programação para esse fim.

Neste caso, uma API (*Application Programming Interface*), em português, Interface de Programação de Aplicações ou Interface de Programação de Aplicação, que consiste em um sistema intermediário que de um lado tem acesso aos dados de banco de dados que contém os metadados e do outro lado pode receber os pedidos de dados provindos da camada Ações do nosso servidor que hospeda o *chatbot*.

Como proposta para o desenvolvimento de uma API para o modelo, a Figura 20 mostra um diagrama contendo o fluxo de trabalho deste módulo.

Figura 20 – Proposta de API



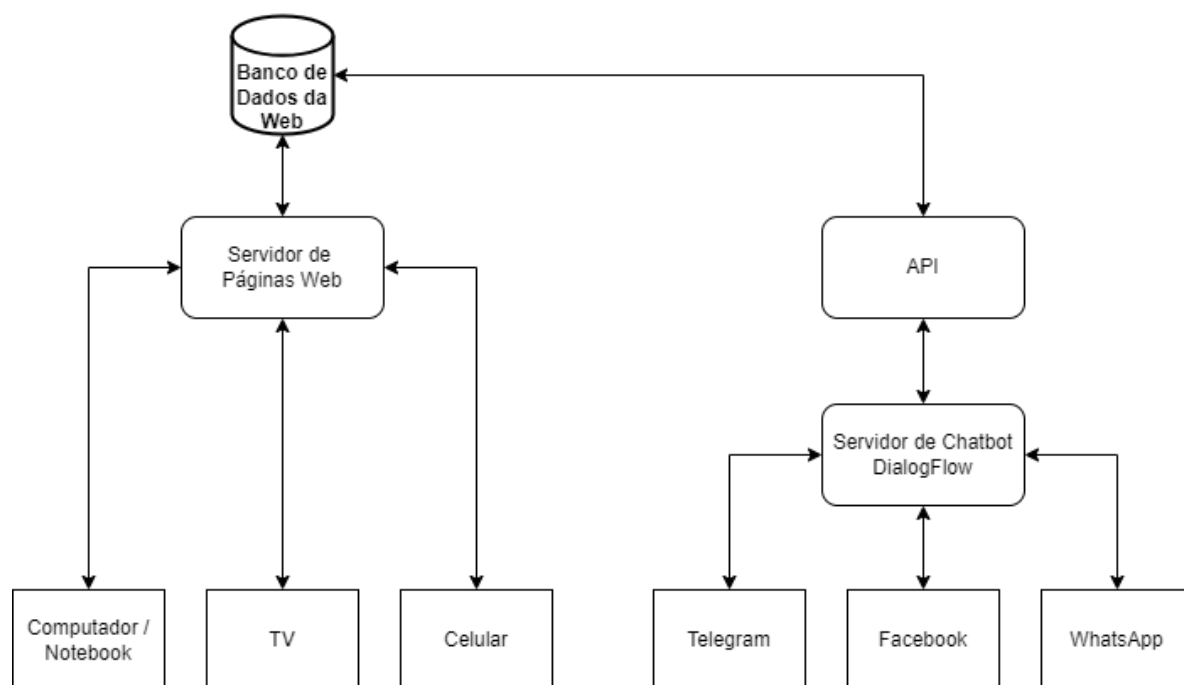
Fonte: elaborado pelo autor

Em sua entrada, a API, recebe a mensagem enviada pelo usuário e direciona aos módulos de reconhecimento de Intenções e de Entidades. Estes termos podem ser trazidos das plataformas de desenvolvimento de *chatbots*, pois os módulos de Diálogo e Ações podem encontrar esses termos e auxiliar no processamento das mensagens. Depois que estes termos foram descobertos, são submetidos ao módulo gerador de expressão de busca que monta uma expressão de busca utilizando-os em conjunto com quaisquer outros termos que estejam armazenados nas variáveis de contexto. Em seguida, é efetuada uma busca para recuperação de informações no banco de dados dos documentos que contém os metadados e as informações encontradas são devolvidas ao módulo gerador. De posse dessas informações, os resultados são enviados ao módulo gerador de respostas que pode construir uma resposta contendo informações a respeito dos documentos, ou metadados, encontrados. Uma cópia de todas as perguntas e suas respectivas respostas são enviadas ao banco de dados que servirá para análise futura, por parte dos desenvolvedores, e avaliação da eficiência do agente e possíveis correções dos módulos envolvidos. Neste momento é apresentado o resultado do questionamento atual ao usuário, e este avaliando pode decidir fazer uma nova pergunta e todo o processo reinicia.

5.1.1 Dialogflow

Ao planejar como os documentos hospedados em um repositório podem ser utilizados por inúmeras interfaces, o desenvolvedor precisa identificar como esse processo deverá ser feito para que no momento da escolha das ferramentas que serão utilizadas, estas tenham a compatibilidade e os recursos necessários para tornar isso possível. Na Figura 21 tem-se uma ilustração de uma plataforma onde os dados estão armazenados em um sistema com banco de dados na Web e dois sistemas diferentes podem consultá-lo de forma a entregar uma experiência bastante rica aos usuários que podem acessar navegadores utilizando por meio de diversos tipos de equipamentos, como computadores, notebook, TVs e celular. Além disso, pode ser usado diversos sistemas de construção de agentes de diálogo, e ao escolher como ferramenta para configurar este servidor do sistema de *chatbot* o produto Dialogflow, por exemplo, com o mínimo de configuração já permite a utilização por inúmeros sistemas conversacionais, como Telegram, Facebook, WhatsApp ou até mesmo em uma janela de *chat* disponibilizada na própria página na Internet.

Figura 21 – Compartilhamento de um mesmo repositório por múltiplas interfaces



Fonte: elaborado pelo autor

O Dialogflow é um desenvolvedor de tecnologias de interação humano-computador com base em conversas em linguagens naturais. Esta ferramenta escolhida para utilização na pesquisa faz parte do pacote de serviços do Google Cloud, que é uma plataforma de soluções

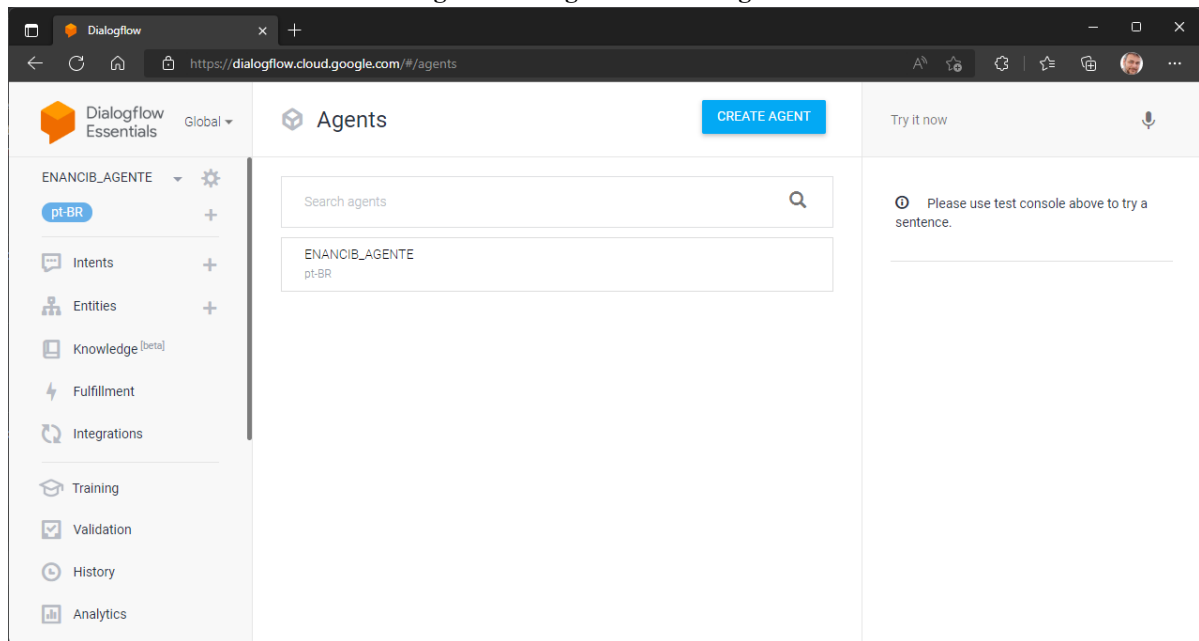
disponibilizadas pela empresa Google e que fornecem produtos e soluções para diversos setores da economia por meio de uma infinidade de aplicativos e serviços destinados a automatizar, transportar para a nuvem ou terceirizar serviços que até pouco tempo seriam implantados em servidores hospedados internamente nas empresas. O produto denominado Dialogflow⁸⁶ faz parte das soluções fornecidas nas áreas de inteligência artificial e aprendizagem de máquina (*machine learning*) e juntamente com outros produtos, fornecem um pacote completo para a criação de sistemas cognitivos e inteligentes. O Dialogflow é uma plataforma de processamento de linguagem natural que facilita o design e a integração de uma interface do usuário conversacional com aplicativos para dispositivos móveis, aplicativos da Web, dispositivos, *bots*, sistemas interativos de resposta de voz etc., e que de acordo com a página do produto, em 2022, existe uma comunidade de mais de 1,5 milhão de desenvolvedores que utilizam essa ferramenta.

Segundo Dias (2020), esta plataforma foi desenvolvida pela Speaktio e lançada em setembro de 2014 como *api.ai*, foi comprada pelo Google em 2016. Em outubro de 2017, foi renomeada de *api.ai* para Dialogflow. E como principais recursos disponibilizados por esta solução é a criação de fluxos visuais que auxiliam o desenvolvimento para que os criadores possam ver, entender, editar e compartilhar seus trabalhos com equipes de desenvolvimento de uma forma mais rápida. Fornece uma inteligência artificial avançada com modelos recentes de PLN capazes de reconhecer as intenções e os contextos com precisão e eficiência, mesmo em casos complexos, pois suporta mais de 30 idiomas e suas variantes.

Um agente conversacional, mostrado na Figura 22, criado pelo sistema do Dialogflow pode processar conversas simultâneas com diversos usuários por meio de vários canais, incluindo entradas de texto ou áudio (como de um *smartphone* ou gravação de voz). Ele também pode responder aos seus clientes de várias maneiras, seja por meio de texto ou com fala sintética. Como possui um módulo de processamento de linguagem natural bastante avançado é capaz de compreender diversas nuances da linguagem humana. Ao receber textos ou áudios de um diálogo, o sistema processa os termos encontrados a um nível em que seus aplicativos ou serviços podem compreender o que está sendo conversado. Além disso com o passar do tempo o sistema aprende com as situações e os diversos cenários que podem acontecer durante os atendimentos e aumenta a sua capacidade de responder a questionamentos cujas informações ele não possui, mas que a experiência pode fornecer a capacidade de inferir sobre os fatos.

⁸⁶ Disponível em: <https://cloud.google.com/dialogflow>. Acesso em: 07 jan. 2022.

Figura 22 – Agentes no Dialogflow

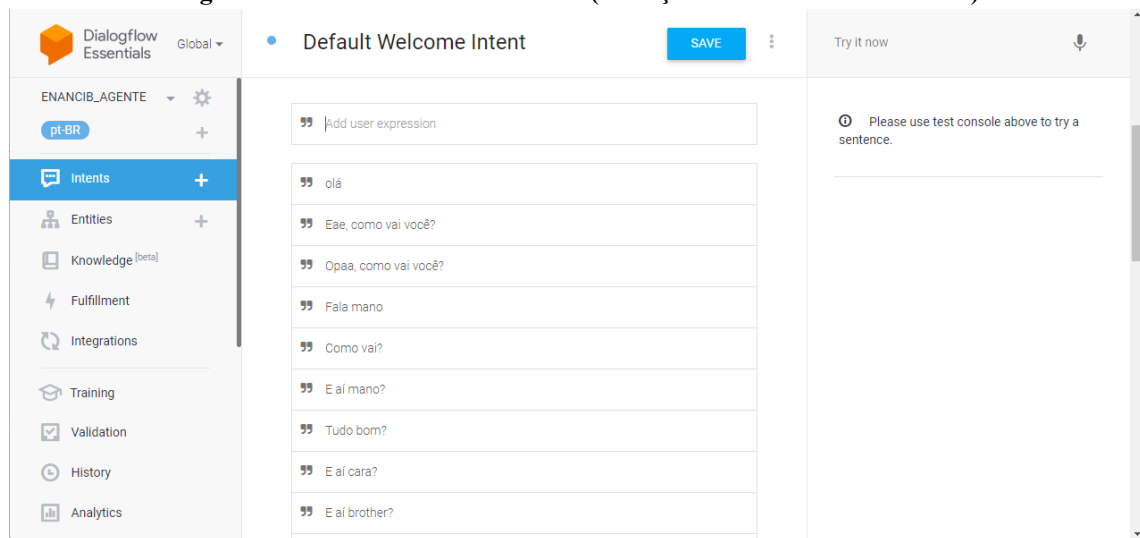


Fonte: elaborado pelo autor

Para compreender como funciona uma interface computacional desse tipo é preciso entender alguns conceitos que fazem parte do processo desenvolvido pela solução na compreensão da linguagem entre o humano e a máquina (DIAS, 2020; GOOGLE, 2022).

Intents - Podem ser traduzidas como intenções de um usuário e servem para mapear os termos que foram inseridos no sistema e a partir daí é possível iniciar um diálogo com o usuário. Quando um usuário final escreve algo, o Dialogflow corresponde o conteúdo à melhor *intent* do seu agente. A correspondência de uma *intent* também é conhecida como classificação de *intent*. Por padrão, ao criar um agente já são criadas duas intenções, uma que serve para controlar as boas-vindas (Figura 23) no início da interação e outra que é acionada todas as vezes em que o sistema não consegue entender o que o usuário disse.

Figura 23 – Default Welcome Intent (Intenção Padrão de Boas-Vindas)

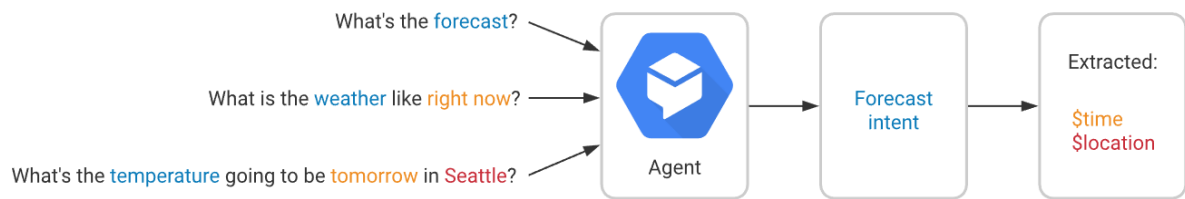


Fonte: elaborado pelo autor

Além desses, outros parâmetros que precisam ser configurados, como os contextos que são utilizados para controlar o fluxo da conversa, os eventos que permitem invocar outras intenções baseado do que acontece no diálogo, as frases de treinamento que o desenvolvedor adiciona para instruir o sistema treinando-o para responder possíveis questionamentos feitos pelos usuários e também quando algo não é compreendido corretamente pelo sistema, podem ser fornecidas novas frases para treinar o sistema de inteligência para que no futuro seja capaz de responder a esse questionamento, ações e parâmetros que auxiliam pequenas interações automatizadas ao consultar bases de dados externas, as respostas do agente que ao invés de haver apenas uma única resposta para um questionamento podem ser adicionados novas respostas, inclusive com a utilização de variáveis, para tornar a comunicação mais humanizada.

Um exemplo da função *intent* na criação de um agente meteorológico, seria necessária uma *intent* para perguntas do usuário sobre a previsão do tempo. Se um usuário perguntasse "Qual é a previsão do tempo?", o Dialogflow corresponderia essa expressão à *intent* de previsão. Também é possível configurar sua *intent* para extrair informações úteis da expressão do usuário final, como um horário ou um local para a previsão do tempo. Esses dados extraídos, juntamente com os dados da pergunta, são importantes para que o sistema execute uma consulta de previsão do tempo para o usuário. A Figura 24 ilustra as várias possibilidades de perguntas do usuário, mas todas culminando em apenas uma *intent*, que extrai os dados necessários e pode prosseguir com a busca das informações.

Figura 24 – A função da *Intent*



Fonte: Disponível em: <https://cloud.google.com/dialogflow/es/docs/basics>. Acesso em: 02 jan. 2022.

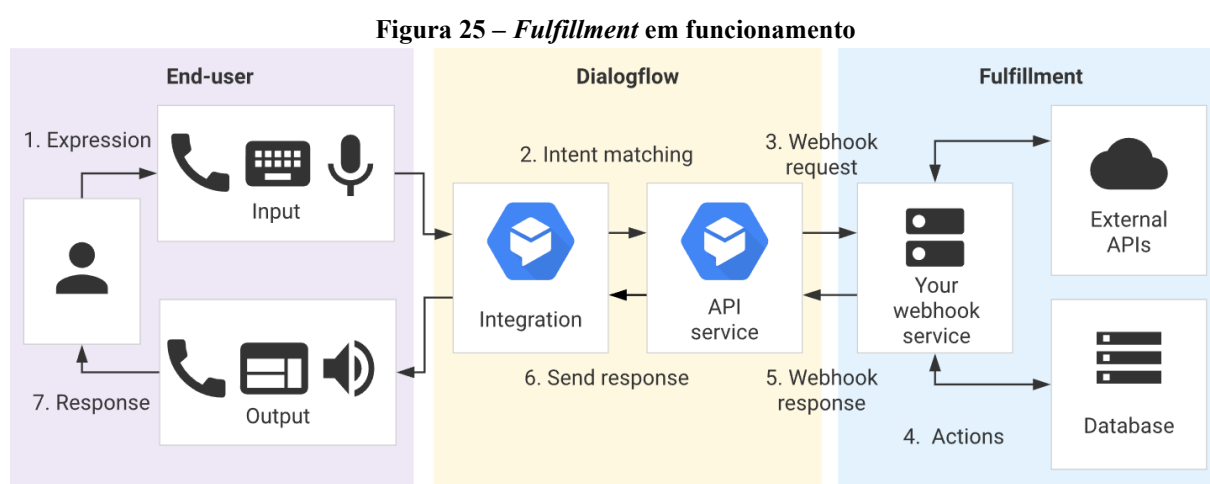
Entities – Podem ser identificadas como entidades e são usadas para criar lista de palavras ou dar um peso à uma palavra específica. Essas entidades podem ser de sistema ou de desenvolvedor. As entidades de sistema são predefinidas e podem ser utilizadas para buscar parâmetros do sistema como datas, horários, cores, endereços de *e-mail* e assim por diante. Já as entidades de desenvolvedor podem ser usadas para qualquer outra informação que seja necessária e não esteja incluída nas entidades do sistema. Aqui são disponibilizadas diversas ferramentas para auxiliar o desenvolvedor no processo de criação das entidades para tratar os termos que foram encontrados e que possam ser usados em seguida. Por exemplo, ao definir uma entidade como vegetal seria possível corresponder todos os tipos de vegetais disponíveis para compra com um agente de mercado.

Knowledge – É um recurso que ao ser habilitado permite que o agente avalie documentos na busca de onde encontrar respostas automatizadas. Ao configurar este módulo, o desenvolvedor define uma ou mais bases de conhecimento, que são coleções de documentos, e a cada solicitação de intenção o sistema pode utilizar essas bases para tentar formular respostas a questionamentos que não estão cadastradas na configuração normal.

Context – Os contextos do Dialogflow são iguais ao contexto da linguagem natural. Por exemplo, se alguém diz que “ele gosta de banana”, é necessário compreender o contexto para se determinar quem é ele. Dessa mesma forma, o Dialogflow precisa processar as expressões dos usuários para atribuir um contexto a elas e conseguir corresponder corretamente a uma intenção. Por meio do controle dos contextos, o sistema consegue controlar o fluxo de uma conversa. Para atribuição de contextos a uma *intent*, é necessário configurar contextos de entrada e saída, identificados por nomes. Quando uma *intent* é correspondida, todos os contextos de saída configurados para essa *intent* são ativados.

Fulfillment – É um código de programação que permite que o agente do sistema chame a lógica de negócio para ser utilizada em uma intenção. Durante as conversas o sistema permite usar as informações extraídas pelo módulo de processamento de linguagem natural para gerar

respostas dinâmicas ou requisitar ações no servidor de dados da empresa. Para que isso seja possível, o desenvolvedor precisa criar uma estrutura de servidor Web e hospedá-la de forma que o sistema possa consultá-la sempre que achar necessário ou estiver configurado para tal em alguma intenção. Quando ocorre uma necessidade dessa interação o sistema cria uma requisição e envia ao servidor que está disponível, que recebe a solicitação e realiza todas as tarefas necessárias, por exemplo para solicitar informações de um produto ou a respeito de um pedido, e em seguida fornece uma resposta com várias informações como o contexto que iniciou a conversa, o nome do evento que provocou esta consulta, os dados requisitados e várias outras informações importantes para a continuação do diálogo. Recomenda-se que o servidor que responde às requisições do *fulfillment* esteja bem configurado e com regras de segurança bem definidas para que pessoas mal-intencionadas não tenham acesso às informações, tentando por meio de requisições falsas, extrair dados deste sistema. A Figura 25 mostra o processo em funcionamento.



Fonte: Disponível em: <https://cloud.google.com/dialogflow/es/docs/basics>. Acesso em: 17 fev. 2022.

Segue a descrição das etapas durante o processo de aquisição de informações de um banco de dados, ou API, externo da solução:

- 1 - O usuário digita ou fala uma expressão interagindo com o sistema.
- 2 - O Dialogflow combina a expressão do usuário a uma intent e extrai parâmetros.
- 3 - O Dialogflow envia uma mensagem de solicitação para o serviço apropriado. Essa mensagem contém informações sobre a intent correspondida, a ação, os parâmetros e a resposta definida para a intent.

4 - O serviço toma as medidas necessárias, como consultar o banco de dados ou chamar um serviço de API externa.

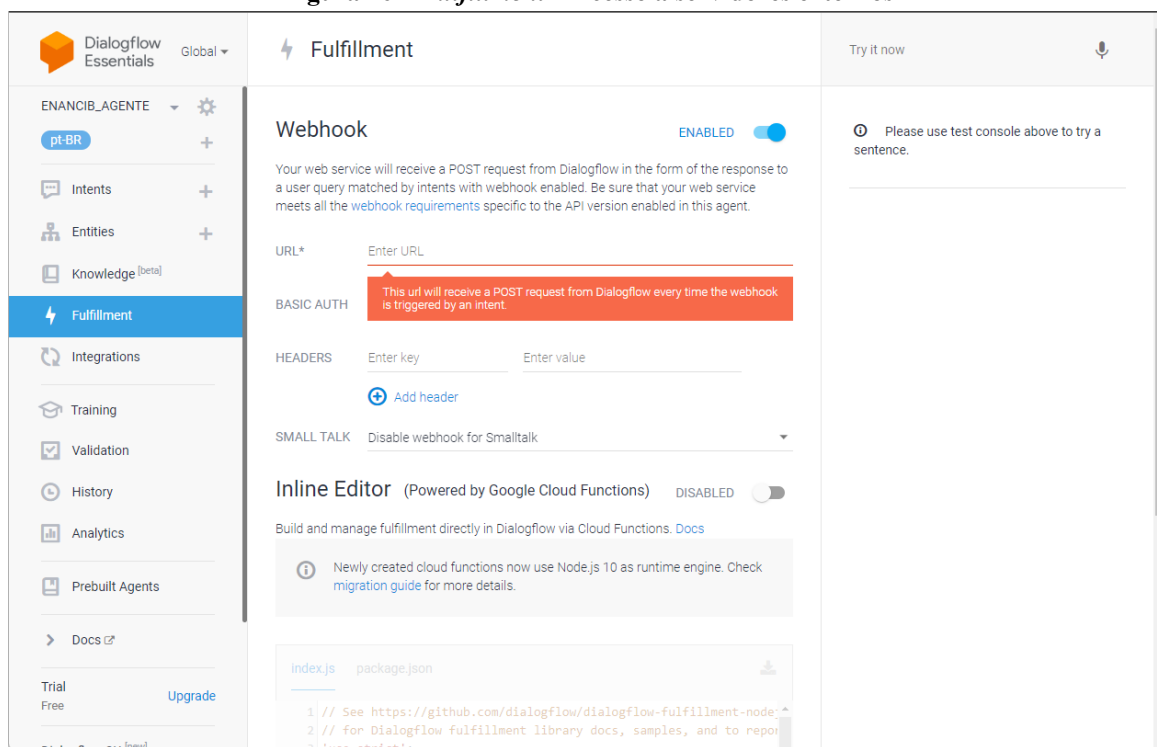
5 - O serviço envia uma mensagem de resposta para o Dialogflow contendo a resposta a ser enviada ao usuário final.

6 - O Dialogflow, por sua vez, faz o que foi solicitado.

7 - O usuário vê ou ouve a resposta.

A Figura 26 mostra a página de configuração de acesso a servidores externos de dados habilitada, pode-se verificar os campos onde são inseridos os dados de integração (HEADERS) e endereços de conexão (URL, BASIC AUTH).

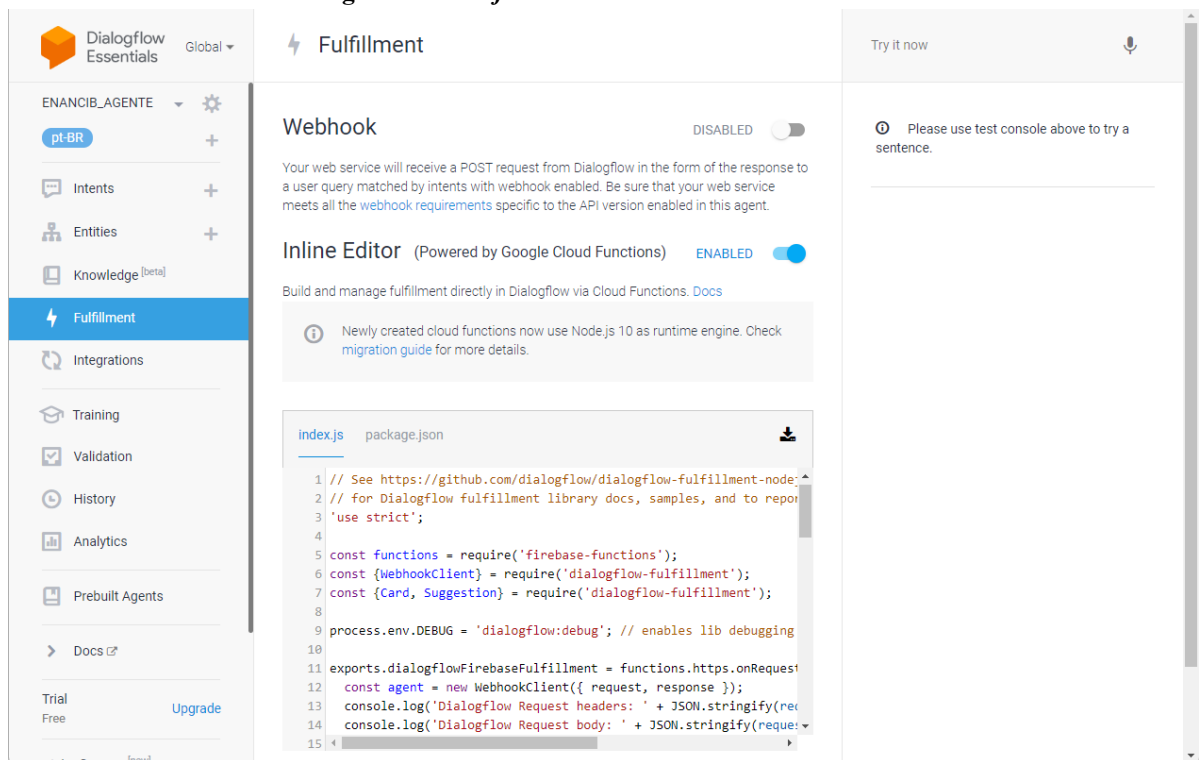
Figura 26 – Fulfillment – Acesso a servidores externos



Fonte: elaborado pelo autor

Uma outra maneira de usar o *Fulfillment* é por meio da plataforma *Cloud Functions* fornecido pelo Google. O editor *in-line* do Dialogflow é integrado ao *Cloud Functions*, conforme Figura 27. Quando você usa o editor *in-line* para criar e editar o código, o Dialogflow estabelece uma conexão segura com a função nos servidores do Google e dessa forma é dispensada a existência de um servidor proprietário ou externo para fornecer os dados.

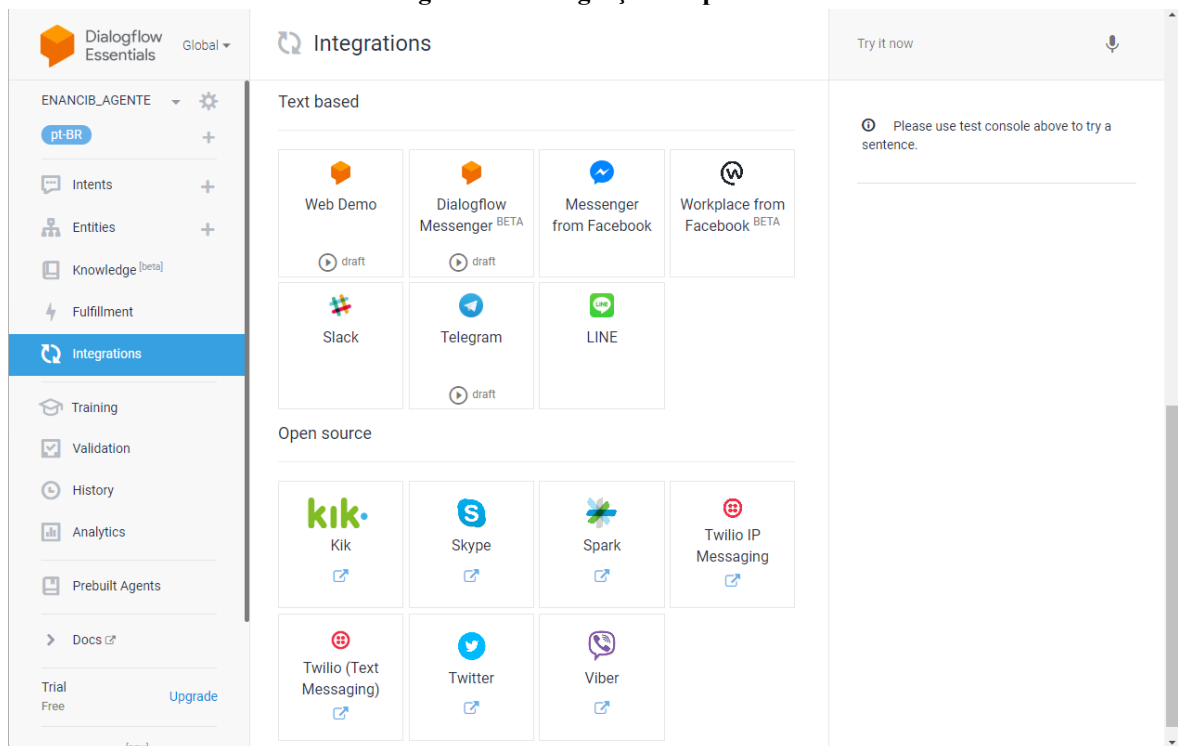
Figura 27 – Fulfillment – Desenvolvimento interno



Fonte: elaborado pelo autor

Integrations – Depois de desenvolver o agente no Dialogflow, é possível por meio de ferramentas de integração, disponibilizá-la em várias plataformas. O módulo de integrações permite colocar o agente criado em diversas plataformas disponíveis de terceiros (Google Assistente, Slack, Facebook Messenger, Twitter, Telegram, Skype, Line, Amazon Alexa, entre outros, além de diversos sistemas de telefonia por voz) para que os usuários possam interagir com ela, ilustrado na Figura 28.

Figura 28 – Integrações disponíveis



Fonte: elaborado pelo autor

Training – O módulo de treinamento permite ao desenvolvedor verificar como o processamento de linguagem natural do Dialogflow, que é baseado em aprendizagem de máquina, está aprendendo, e por meio de relatórios, fornece uma interface para que possa ser incorporado registros externos, internos, ou de interação com clientes com base em frases utilizadas pelos atendentes humanos. Outra função é a possibilidade de exportar esse recurso de treinamento de um agente em funcionamento para um novo agente e melhorar assim o desempenho desde o início do processo. Uma tela onde é possível verificar a lista das conversas que foram efetuadas entre os clientes e o agente, lá um analista pode analisar e trabalhar para que a cada dia o sistema possa ser aprimorado, para consequentemente melhorar o desempenho no processo de respostas aos questionamentos, o que aumenta o grau de relevância das respostas e diminui o número de solicitações sem correspondência, pelo motivo do sistema não ter encontrado a intenção do usuário.

Validation – Esse recurso de validação permite ter acesso aos resultados que são automaticamente criados sempre que o treinamento do agente é realizado e concluído. Estes resultados fornecem uma lista de erros e avisos que devem ser corrigidos para melhorar a qualidade e o desempenho do agente.

History – Nesta página é mostrado uma versão simplificada das conversas com o agente. São registros cronológicos e proporciona uma visão de como os usuários interagem com ele, sendo possível emitir relatórios individuais de cada mensagem para uma análise mais aprofundada.

Analytics – Esta é uma das páginas mais importantes na busca da melhoria dos serviços do agente, pois ela oferece uma série de informações a respeito do desempenho dele. Por ser uma ferramenta de criação de agentes bastante dinâmica, onde é capaz de aprender com as interações com os usuários, é importante que estas sejam avaliadas através de métricas bem definidas na busca da qualidade dos resultados que são entregues aos usuários, informações como o número de sessões e consultas por sessão, ou mesmo, relatórios detalhados a respeito das intenções mais utilizadas e sua taxa de acerto, relatórios do módulo de processamento de linguagem natural, tudo isso destinado para análise posterior e futuras correções.

Prebuilt Agents – O sistema fornece, por meio desta janela, uma coleção de agentes desenvolvidos pela própria equipe do Dialogflow com o objetivo de auxiliar desenvolvedores iniciantes na construção de seus agentes ou na instrução de novos usuários, pois é possível a partir destas coleções realizar alterações e criar agentes que poderão ser utilizados em suas empresas.

5.1.2 ENANCIB AGENTE – *Chatbot conceitual*

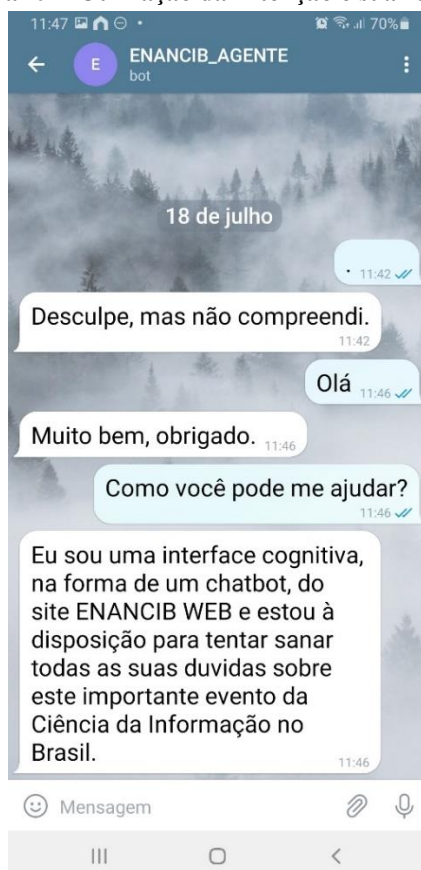
Com base no que foi descrito até aqui, foi concebido o modelo de um sistema conversacional para servir de interface cognitiva para o sítio ENANCIB WEB, que fornece acesso aos trabalhos apresentados nas edições do Encontro Nacional de Pesquisa e Pós-graduação em Ciência da Informação, hospedado em servidores Windows na Web, desenvolvido utilizando a plataforma ASP.NET e um banco de dados hospedado em servidores Linux na Web utilizando o sistema de gerenciamento de banco de dados MySQL; e como solução para desenvolvimento de agentes foi escolhido o produto Dialogflow fornecido pela empresa Google.

Para simular a utilização da ferramenta e testar o conceito do modelo, durante o processo de diálogo, existe a necessidade do acompanhamento e análise dos termos fornecidos pelo usuário, para que sejam encontrados alguns parâmetros descritos como fundamentais para a integração entre as ferramentas, que são as **Intenções**, **Entidades** e **Contextos**. A partir dessas informações é possível iniciar a construção dos comandos de busca que serão aplicados sobre

a base de dados para extraí-los, e seus metadados. O processo deve prosseguir enquanto houver respostas a serem enviadas ao usuário com informações a respeito dos documentos armazenados, para a construção das respostas e a continuidade do diálogo até a necessidade do usuário ser satisfeita ou ele ser informado que não será possível prosseguir utilizando este meio de acesso.

Em nosso exemplo, todas as **Intenções** podem ser classificadas como desejos do usuário, é quando ele expressa a sua vontade por meio de um termo que representa o tipo de documento ou informação que ele deseja que o sistema recupere para ele. A Figura 29 mostra um exemplo da aplicação deste processo quando o sistema detecta uma intenção de **AJUDA** por meio da palavra **AJUDAR** e depois de extraí-la e processá-la no modelo é devolvida a seguinte resposta **“Eu sou uma interface cognitiva, na forma de um *chatbot*, do site ENANCIB WEB e estou à disposição para tentar sanar todas as suas dúvidas sobre este importante evento da Ciência da Informação no Brasil.”** que atende aos anseios do usuário por conter informações relevantes, visto que ele não foi claro sobre o que precisava de ajuda por não fornecer nenhuma outra informação por meio de termos adjacentes.

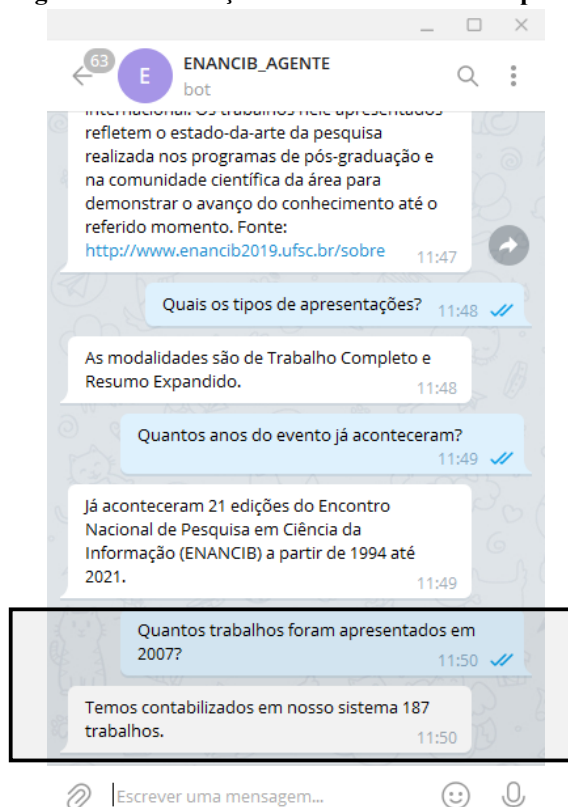
Figura 29 – Utilização da Intenção e sua resposta



Fonte: elaborado pelo autor

As **Entidades** são palavras que provavelmente estarão entre os termos dos documentos e seus metadados, são palavras que poderão ser encontradas nos títulos, autores, palavras-chave, resumo, etc. É por meio deste termo que o sistema é traduzido para o formato de uma variável e aplicado à expressão de busca para servir como uma palavra-chave da intenção. Neste momento que a riqueza dos metadados a respeito dos documentos farão a diferença e é o mote desta pesquisa, pois a existência de muitos metadados devidamente etiquetados é que pode facilitar o processo de encontrar o termo a ser recuperado e por meio das ligações entre eles, enriquecer o diálogo com o usuário. A Figura 30 mostra um caso em que a pergunta do usuário **“Quantos trabalhos foram apresentados em 2007?”** foram encontradas as seguintes entidades: **TRABALHOS** e **2007**, além da intenção **QUANTOS** que significa a necessidade de sumarizar os dados, neste momento foi construída uma requisição ao banco de dados que continham as duas variáveis preenchidas, o tipo de documento: **TRABALHOS** e um ano de edição: **2007**. Depois de efetuada a busca dos dados é gerada uma resposta com os resultados, mas como uma grande quantidade de documentos foram encontrados e seu desejo era de ser sumarizado, o sistema apresenta apenas a somatória com a frase **“Temos contabilizados em nosso sistema 187 trabalhos.”** e fica aguardando mais perguntas para prosseguir com o refino das respostas.

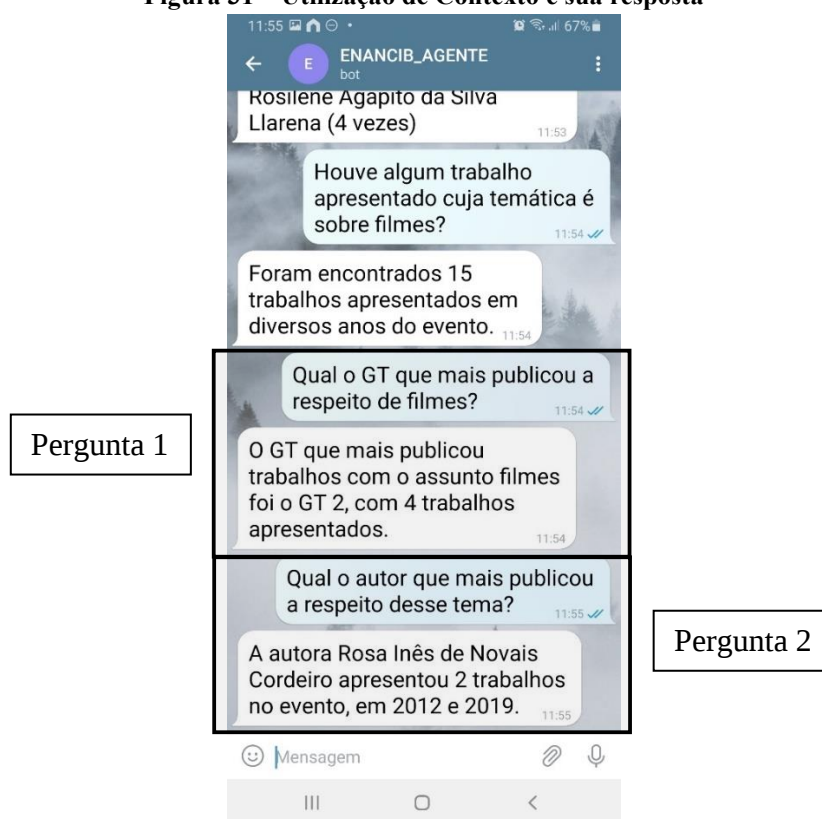
Figura 30 – Utilização de Entidades e sua resposta



Fonte: elaborado pelo autor

Já o **Contexto** é necessário na comunicação porque dentro de um mesmo assunto é permitido que diversos diálogos sejam trocados sem a utilização de todos os termos necessários para demonstrar que o assunto ainda não foi alterado. Como por exemplo, o usuário cita um ano e depois começa a criar perguntas a respeito dos documentos, não voltando a citar o ano, dessa forma o contexto de todos os questionamentos continua sendo o ano citado até que o usuário mude essa informação e comece um novo ciclo de assunto durante o diálogo. A Figura 31 demonstra esse processo, ao fazer o seguinte questionamento, mostrado na pergunta 1, com os termos “**Qual o GT que mais publicou a respeito de filmes?**”, a partir da **Intenção: QUAL**, foram encontrados as seguintes **Entidades: GT** e **filmes**, onde o primeiro se refere a um metadado de qual grupo de trabalho pertence os documentos e o segundo se refere a um termo que pode ser encontrado nos metadados de títulos, palavras-chave e resumo e pode indicar o assunto a que se refere o questionamento. Após efetuar a consulta no banco de dados e receber as informações, o sistema sumariza as quantidades de trabalhos encontrados por grupo de trabalho e apresenta o resultado, ao usuário, de apenas um grupo, o que continha a maior quantidade de trabalhos, o GT 2, respondido da seguinte forma: “**O GT que mais publicou trabalhos com o assunto filmes foi o GT 2, com 4 trabalhos apresentados.**”.

Figura 31 – Utilização de Contexto e sua resposta



Fonte: elaborado pelo autor

A pergunta 2 é solicitada utilizando o seguinte conjunto de termos “**Qual o autor que mais publicou a respeito desse tema?**”, ao analisar os termos encontram-se a **Intenção: QUAL** e as **Entidades: AUTOR** e **TEMA**, e como o assunto da pergunta anterior se refere a um tema de trabalho e neste momento não foram fornecidos novos atributos para nortear alguma mudança de assunto, o sistema infere que o tema continua sendo o assunto da pergunta anterior, que versava sobre **FILMES**, e partindo dessa premissa, acoplando a este termo herdado o termo encontrado de **AUTOR**, pode se formular uma nova expressão de busca e efetuar uma nova consulta ao banco de dados para investigar a existência de autores que publicaram sobre filmes, sumarizando os resultados e apresentando apenas o autor com o maior número, ou em caso de empate, todos os autores empatados em primeiro lugar. A resposta apresentada foi a seguinte “**A autora Rosa Inês de Novais Cordeiro apresentou 2 trabalhos no evento, em 2012 e 2019.**”.

No desenvolvimento de um ambiente para a simulação do funcionamento do modelo, foi criado um texto com determinadas frases no formato de questionamentos fictícios e aleatórios a respeito da temática contida nos trabalhos publicados no evento, compilados e disponibilizados por meio da plataforma ENANCIB WEB, para que a partir de questionamentos

provindos de usuários, por meio de diferentes interfaces conversacionais, os processos de análise e identificação de intenção, entidades e contextos pudessem ser executados, então, a partir dos termos selecionados as buscas fossem efetuadas na base de dados do portal ENANCIB WEB, o resultado permitiria a utilização dos metadados para construção das respostas e apresentação ao usuário. Esse processo deve se repetir enquanto acontecer o diálogo ou até o usuário ficar satisfeito, repetindo:

- Usuário faz uma pergunta,
- Sistema recebe a pergunta,
- Separa os termos de intenção, entidades e contextos,
- Formula a questão de consulta,
- Aplica a consulta ao banco,
- Recebe os resultados da consulta ao banco,
- Processa os resultados por meio de Aplicação Externa,
- Transforma estes resultados para linguagem natural e
- Apresenta ao usuário.

O Quadro 10 apresenta um longo diálogo que ocorreu entre um usuário e o sistema do modelo, a coluna 1 apresenta os pares de perguntas e respostas, pois sempre que o usuário fizer uma pergunta ele recebe uma resposta logo em seguida, então são feitas novas perguntas e novas respostas, e assim por diante. A coluna 2 identifica o personagem do diálogo, que se alterna entre usuário e sistema. Na coluna 3 é mostrado o enunciado da pergunta e o texto gerado pelo sistema e apresentado ao usuário como uma resposta. E na última coluna chamada “ato interno do sistema” ela apresenta quais foram os termos encontrados nas perguntas do usuário e os classifica para ser utilizado no processo de recuperação das informações e o seu resultado é apresentado no campo seguinte do quadro.

Quadro 10 – Diálogo entre usuário e *chatbot* do Projeto

Interação	Personagem	Enunciado	Ato Interno do sistema
1	Usuário	Olá	Identifica um cumprimento (olá)
1	Sistema	Muito bem, obrigado.	
2	Usuário	Como você pode me ajudar?	Intenção encontrada: ajudar
2	Sistema	Eu sou uma interface cognitiva, na forma de um chatbot, do site ENANCIB WEB e estou à disposição para tentar sanar todas as	

		suas dúvidas sobre este importante evento da Ciência da Informação no Brasil.	
3	Usuário	Você pode me falar mais sobre este evento?	Intenção encontrada: falar Entidades encontradas: sobre evento
3	Sistema	O Encontro Nacional de Pesquisa em Ciência da Informação (ENANCIB) é o principal evento de pesquisa e de pós-graduação da área de Ciência da Informação do País e visa discutir e refletir a produção de conhecimento na área, de modo a estimular, por meio de amplo diálogo entre os pesquisadores que nela atuam, a realidade dos programas de pós-graduação. Consiste em um foro privilegiado para a reflexão e o compartilhamento das pesquisas científicas em Ciência da Informação e áreas de interface, congregando pesquisadores, docentes e pós-graduandos. O evento é voltado à troca de experiências acadêmico-científicas e pelo fortalecimento de laços acadêmicos em nível nacional e internacional. Os trabalhos nele apresentados refletem o estado-da-arte da pesquisa realizada nos programas de pós-graduação e na comunidade científica da área para demonstrar o avanço do conhecimento até o referido momento. Fonte: http://www.enancib2019.ufsc.br/sobre	
4	Usuário	Quais os tipos de apresentações?	Intenção encontrada: quais Entidades encontradas: tipos apresentações
4	Sistema	As modalidades são de Trabalho Completo e Resumo Expandido.	
5	Usuário	Quantos anos do evento já aconteceram?	Intenção encontrada: quantos Entidades encontradas: anos evento
5	Sistema	Já aconteceram 21 edições do Encontro Nacional de Pesquisa em Ciência da Informação (ENANCIB) a partir de 1994 até 2021.	
6	Usuário	Quantos trabalhos foram apresentados em 2007?	Intenção encontrada: quantos Entidades encontradas: trabalhos 2007

6	Sistema	Temos contabilizados em nosso sistema 187 trabalhos.	
7	Usuário	E em 2021?	Entidades encontradas: 2021 Contexto: trabalhos
7	Sistema	Em 2021 foram apresentados 342 trabalhos no evento.	
8	Usuário	Em 2021 como se dividiu o evento?	Entidades encontradas: 2021 evento dividiu
8	Sistema	Em todos os anos do evento, os trabalhos são divididos em Grupos de Trabalhos (GTs) de acordo com os assuntos abordados e os temas das apresentações.	
9	Usuário	Quais os grupos de trabalhos de 2021?	Intenção encontrada: quais Entidades encontradas: grupos de trabalho 2021
9	Sistema	Em 2021 foram divididos em GTs da seguinte forma: GT 1 – Estudos Históricos e Epistemológicos da Ciência da Informação GT 2 – Organização e Representação do Conhecimento GT 3 – Mediação, Circulação e Apropriação da Informação GT 4 – Gestão da Informação e do Conhecimento GT 5 – Política e Economia da Informação GT 6 – Informação, Educação e Trabalho GT 7 – Produção e Comunicação da Informação em Ciência, Tecnologia & Inovação GT 8 – Informação e Tecnologia GT 9 – Museu, Patrimônio e Informação GT 10 – Informação e Memória GT 11 – Informação & Saúde	
10	Usuário	Quantos trabalhos foram apresentados no GT 8?	Intenção encontrada: quantos Entidades encontradas: trabalhos GT 8
10	Sistema	Foram apresentados no GT 8 - Informação e Tecnologia um total de 43 trabalhos.	
11	Usuário	Quais os assuntos das palavras-chave mais apresentados em 2021?	Intenção encontrada: quais Entidades encontradas: assuntos

			palavras-chave 2021
11	Sistema	Os 10 assuntos mais retratados pelas palavras-chave em 2021 foram os seguintes: Ciência da Informação (33 vezes) gestão da informação (17 vezes) Organização do Conhecimento (13 vezes) competência em informação (12 vezes) informação (11 vezes) Mediação da Informação (11 vezes) memória (11 vezes) Produção científica (11 vezes) Agenda 2030 (9 vezes) Metadados (9 vezes)	
12	Usuário	Quais os autores que mais apresentaram trabalhos em 2021?	Intenção encontrada: quais Entidades encontradas: autores trabalhos 2021
12	Sistema	Os autores que mais apresentaram em 2021 foram os seguintes: Dalton Lopes Martins (5 vezes) Gustavo Silva Saldanha (5 vezes) Silvana Aparecida Borsetti Gregório Vidotti (5 vezes) Alzira Karla Araújo da Silva (4 vezes) Carlos Cândido de Almeida (4 vezes) Cláudio Paixão Anastácio de Paula (4 vezes) Deise Maria Antonio Sabbag (4 vezes) Dirnele Carneiro Garcez (4 vezes) Elisabete Gonçalves de Souza (4 vezes) Ieda Pelógia Martins Damian (4 vezes) Izabel França de Lima (4 vezes) Márcia Cristina de Carvalho Pazin Vitoriano (4 vezes) Maria Luíza de Almeida Campos (4 vezes) Marli Dias de souza Pinto (4 vezes) Marta Lúcia Pomim Valentim (4 vezes) Rosilene Agapito da Silva Llerena (4 vezes)	
13	Usuário	Houve algum trabalho apresentado cuja temática é sobre filmes?	Intenção encontrada: houve Entidades encontradas: trabalho filmes
13	Sistema	Foram encontrados 15 trabalhos apresentados em diversos anos do evento.	

14	Usuário	Qual o GT que mais publicou a respeito de filmes?	Intenção encontrada: qual Entidades encontradas: GT filmes
14	Sistema	O GT que mais publicou trabalhos com o assunto filmes foi o GT 2, com 4 trabalhos apresentados.	
15	Usuário	Qual o autor que mais publicou a respeito desse tema?	Intenção encontrada: qual Entidades encontradas: autor tema Contexto: filmes
15	Sistema	A autora Rosa Inês de Novais Cordeiro apresentou 2 trabalhos no evento, em 2012 e 2019.	
16	Usuário	Quantos trabalhos o GT 2 publicou do início até hoje?	Intenção encontrada: quantos Entidades encontradas: GT 2 início hoje
16	Sistema	Foram apresentados no GT 2 um total de 719 trabalhos de 1994 a 2021.	
17	Usuário	Quantos GTs havia no primeiro ano do evento?	Intenção encontrada: quantos Entidades encontradas: GTs primeiro ano evento
17	Sistema	Em 1994, o primeiro ano do evento havia 7 grupos de trabalho: Informação Tecnológica Informação e Sociedade/Ação Cultural Representação do Conhecimento/Indexação/Teoria da Classificação Administração/Gestão/Avaliação e estudos de usuário Formação profissional/Mercado de trabalho Produção científica/Literatura cinza Políticas de Pesquisa dos Cursos de Pós-Graduação	

18	Usuário	Quem foi o autor que mais publicou a respeito de tesouros?	Intenção encontrada: quem Entidades encontradas: autor tesouros
18	Sistema	Foram encontradas 30 apresentações e os autores que mais apresentaram a respeito do assunto tesouro foram: Maria Luiza de Almeida Campos (4 vezes) Mariângela Spotti Lopes Fujita (4 vezes) Benildes Coura Moreira dos Santos Maculan (3 vezes) Brígida Maria Nogueira Cervantes (3 vezes)	
19	Usuário	E quem foi o autor que mais publicou a respeito de recuperação de informação?	Intenção encontrada: quem Entidades encontradas: autor recuperação de informação
19	Sistema	Foram encontradas 49 apresentações e os autores que mais apresentaram a respeito do assunto recuperação de informação foram: Renata Maria Abrantes Baracho (5 vezes) Beatriz Valadares Cendon (4 vezes) Claudio Gottschalg Duque (3 vezes) Marisa Brascher (3 vezes)	
20	Usuário	Voltando a 2021, houve algum trabalho sobre a Amazônia?	Intenção encontrada: houve Entidades encontradas: 2021 trabalho Amazônia
20	Sistema	Foram encontrados 9 trabalhos com a temática ligada à Amazônia, mas nenhum trabalho em 2021.	
21	Usuário	E sobre natureza?	Entidades encontradas: natureza Contexto: 2021 trabalho
21	Sistema	Foram encontrados 8 trabalhos com a temática ligada à natureza, e 1 trabalho no ano de 2021.	

22	Usuário	Me mostre este trabalho.	Intenção encontrada: mostrar Contexto: 2021 trabalho natureza
22	Sistema	Ano: 2021 GT 1 - Estudos Históricos e Epistemológicos da Ciência da Informação Título: Controvérsias Científicas da Natureza Interdisciplinar da Ciência da Informação: Rastros e Inscrições de Cientistas da Informação Palavras-Chave: interdisciplinaridade Ciência da Informação natureza interdisciplinar controvérsias científicas Autores: Zayr Claudio Gomes da Silva Marlene Oliveira Link: https://enancib.ancib.org/index.php/enancib/xxienancib/paper/view/94/139	
23	Usuário	Mudando de assunto, existem trabalhos sobre inteligência artificial nesse evento?	Intenção encontrada: existem Entidades encontradas: trabalhos inteligência artificial evento
23	Sistema	Foram encontrados 5 trabalhos apresentados em diversos anos do evento.	
24	Usuário	Me mostre os títulos.	Intenção encontrada: mostrar Contexto: trabalhos inteligência artificial
24	Sistema	Os 5 trabalhos com a temática sobre inteligência artificial encontrados são: 1 - INTELIGÊNCIA ARTIFICIAL E CONTROLE SOCIAL DA	

		CT&I: UMA RELAÇÃO PERTINENTE À CIÊNCIA DA INFORMAÇÃO (2014) 2 - Análise da produção científica em Inteligência Artificial na área da Ciência da Informação no Brasil (2018) 3 - INTELIGÊNCIA ARTIFICIAL EM CIÊNCIA DA INFORMAÇÃO: REVISÃO SISTEMÁTICA DA LITERATURA (2021) 4 - INTELIGÊNCIA ARTIFICIAL E CIÊNCIA DE DADOS EM CRIS INSTITUCIONAL: MODELO CONCEITUAL (2021) 5 - INTELIGÊNCIA ARTIFICIAL NA TUTORIA EM EAD: PERSPECTIVAS NA CIÊNCIA DA INFORMAÇÃO (2021)	
25	Usuário	Me mostre o trabalho 3	Intenção encontrada: mostrar Entidade encontrada: trabalho 3 Contexto: trabalhos inteligência artificial
25	Sistema	Ano: 2021 GT 8 - Informação e Tecnologia Título: INTELIGÊNCIA ARTIFICIAL EM CIÊNCIA DA INFORMAÇÃO: REVISÃO SISTEMÁTICA DA LITERATURA Palavras-Chave: inteligência artificial pós-humanismo Ciência da Informação Autores: Adelaide Helena Targino Casimiro Wagner Junqueira de Araújo Link: https://enancib.ancib.org/index.php/enancib/xxienancib/paper/view/8/442	

26	Usuário	Estes autores já publicaram em outros anos do evento?	Intenção encontrada: publicaram Entidade encontrada: autores anos Contexto: trabalhos inteligência artificial
26	Sistema	Outras publicações destes autores são: Adelaide Helena Targino Casimiro 2018 - GESTÃO DO CONHECIMENTO E PRÁTICAS ARQUIVÍSTICAS: CONSTRUTOS TEÓRICOS PARA FORMAÇÃO E PRESERVAÇÃO DA MEMÓRIA ORGANIZACIONAL 2021 - MEDIAÇÃO DA INFORMAÇÃO E COMPETÊNCIA EM INFORMAÇÃO: UMA RELAÇÃO POSSÍVEL EM CENÁRIO PANDÊMICO 2021 - INTELIGÊNCIA ARTIFICIAL EM CIÊNCIA DA INFORMAÇÃO: REVISÃO SISTEMÁTICA DA LITERATURA Wagner Junqueira de Araújo 2019 - PROSPECÇÃO DE CENÁRIOS: ESTUDO SOBRE A IMPLANTAÇÃO DO PROCESSO ELETRÔNICO NACIONAL NA UNIVERSIDADE FEDERAL DA PARAÍBA 2018 - VALIDAÇÃO DO MAPEAMENTO DE FLUXOS DE INFORMAÇÃO EM PROCESSOS ORGANIZACIONAIS: UMA ABORDAGEM COM FOCO ARQUIVÍSTICO 2018 - CIBER TERRORISMO NA PARAÍBA 2017 - ESTUDO DE CENÁRIOS PROSPECTIVOS PARA O MERCADO EDITORIAL DO LIVRO DIGITAL NO BRASIL 2017 - SITES DOS MUNICÍPIOS DA PARAÍBA: ANÁLISE DE VULNERABILIDADES COMPUTACIONAIS 2016 - ANÁLISE DA DIMENSÃO HUMANA NO PROCESSO DE GESTÃO DE SEGURANÇA DA INFORMAÇÃO 2016 - APLICAÇÃO DO OPEN MONOGRAPH PRESS POR EDITORAS BRASILEIRAS 2015 - ASPECTOS HUMANOS NA GESTÃO DA SEGURANÇA DA INFORMAÇÃO 2015 - Aspectos de Vulnerabilidades em Bibliotecas Digitais Acessíveis 2014 - ANÁLISE DO SISTEMA DE INFORMAÇÃO DA BIBLIOTECA CENTRAL DA UFPB NOS PROCESSOS DE GESTÃO DA INFORMAÇÃO PARA O SETOR DE REFERÊNCIA 2013 - ANÁLISE DE INFORMAÇÕES PESSOAIS NA WEB: métrica para identificar o grau de exposição da informação 2013 - ANÁLISE DE RISCO NO SISTEMA DE CONCESSÃO DE DIÁRIAS E PASSAGENS (SCDP): estudo de caso sob a ótica da segurança da informação no departamento contábil da UFPB 2010 - A segurança do conhecimento nas práticas da gestão da	

		segurança da informação e da gestão do conhecimento 2010 - Relações de colaboração entre os programas de pós-graduação stricto sensu brasileiros na área da ciência da informação: modelagem baseada em grafos e programa de informetria 2021 - PERSPECTIVAS DA PUBLICAÇÃO DE LIVROS EM FORMATO DIGITAL PELAS EDITORAS UNIVERSITÁRIAS NA ÁMERICA LATINA 2021 - INTELIGÊNCIA ARTIFICIAL EM CIÊNCIA DA INFORMAÇÃO: REVISÃO SISTEMÁTICA DA LITERATURA	
27	Usuário	Obrigado pela ajuda, Bye Bye	Intenção encontrada: Despedida Entidade encontrada: Obrigado Bye Bye
27	Sistema	Agradeço a sua visita e espero que eu tenha ajudado. Sempre que precisar estarei aqui à disposição. Até mais e tenha uma Boa Semana.	Finaliza a sessão do diálogo

Fonte: elaborado pelo autor

Analisando o texto com as frases do diálogo, é possível identificar diversos termos que serão utilizados durante o processo de enriquecimento semântico no modelo deste trabalho. O Quadro 11 apresenta todas as intenções que foram capturadas durante o processo de diálogo, além disso são evidenciados os significados que foram atribuídos aos termos para que pudessem ser utilizados na recuperação de informações a respeito dos documentos armazenados no repositório do projeto.

Quadro 11 – Intenções

Intenção encontrada:	Significado
ajudar	O usuário deseja receber informações sobre algum assunto ou documento.
Cumprimento	É feito um cumprimento formal ao sistema de <i>chatbot</i> .
Despedida	Ele despede-se e depois o sistema encerra a sessão do diálogo.
existem	Pergunta sobre a existência ou não de algum parâmetro nos trabalhos.
falar	Pede que o sistema informe sobre algum assunto ou documento.
houve	Pergunta sobre a existência ou não de algum parâmetro nos trabalhos.
mostrar	Pede que o sistema mostre todas as informações de um trabalho específico.
publicaram	Pergunta sobre autores que efetuaram publicações em edições do evento.
quais	Requer informações sobre autores ou trabalhos específicos.
qual	Requer informações sobre autor ou trabalho específico.
quantos	Pede uma contagem de trabalhos sobre um determinado assunto ou autor.
quem	Requer informações sobre o autor de um trabalho ou assunto específico.

Fonte: elaborado pelo autor

A maioria das intenções encontradas são provindas de termos que expressam os desejos do usuário do agente, pois além de expressar sentimentos como cumprimento e despedida, ele diz ao agente se deseja uma quantificação, a existência, informações sobre, qualificação, propriedade, etc. Em conjunto com outros termos provenientes dos metadados, é possível estabelecer um diálogo que possui uma qualidade superior devido à ligação entre os termos e sobre a origem dos dados que foram adicionados ao sistema, muitos provenientes de vocabulários controlados.

Prosseguindo, a aplicação conseguiu identificar quais os desígnios do usuário, para que possa ser possível o prosseguimento, e conseguinte recuperação das informações, mais termos são necessários, neste caso as entidades. O Quadro 12 apresenta todas as entidades que foram extraídas pelo modelo no texto apresentado.

Quadro 12 – Entidades

Entidades encontradas:	Significado
2007	Ano que deseja o trabalho ou trabalho sobre algo desse ano.
2021	Ano que deseja o trabalho ou trabalho sobre algo desse ano.
3	Quantidade específica de algo ou código de algo.
amazônia	Local específico em algum trabalho ou trabalho sobre este local.
ano	Refere-se ao dado de um ano.
anos	Refere-se a um intervalo de anos ou anos específicos.
apresentações	Determina o tipo de trabalho desejado.
assuntos	Determina sobre o que está requerendo ou mais de um.
autor	Determina a autoria de um trabalho ou trabalhos.
autores	Determina as autorias de um trabalho ou trabalhos.

Bye Bye	Expressão de despedida.
dividiu	Forma de como o evento pode ser organizado.
evento	Informações sobre a edição ou edições do evento referenciado.
filmes	Um tipo de assunto que pode ser abordado nos trabalhos.
grupos de trabalho	Forma de como o evento pode ser dividido e organizado.
GT	Forma de como o evento pode ser dividido e organizado.
GT 2	Uma das classificações de como o evento é dividido e cuja temática dos trabalhos semelhantes são agrupados.
GT 8	Uma das classificações de como o evento é dividido e cuja temática dos trabalhos semelhantes são agrupados.
GTs	Conjunto de divisões de como o evento pode ser dividido e organizado.
hoje	Período ou data que se refere.
início	Refere-se a data inicial de alguma classificação.
inteligência artificial	Um tipo de assunto que pode ser abordado nos trabalhos.
natureza	Um tipo de assunto que pode ser abordado nos trabalhos.
Obrigado	Expressão de agradecimento e despedida.
palavras-chave	Metadado referente a trabalhos do evento.
primeiro	Refere-se a posição inicial de alguma classificação.
recuperação de informação	Um tipo de assunto que pode ser abordado nos trabalhos.
sobre	Determina as informações a respeito de algum fato ou documento.
tema	Determina sobre o assunto que está requerendo informações.
tesauros	Um tipo de assunto que pode ser abordado nos trabalhos.
tipos	Variedades de informações sobre algum tipo de classificação.
trabalho	Determina informações a respeito de um trabalho específico.
trabalhos	Determina informações a respeito de trabalhos específicos.

Fonte: elaborado pelo autor

Estes termos possuem grande importância quando ocorre o processo de recuperação das informações que estão armazenadas no repositório do banco de dados. A maioria das palavras encontradas sob a etiqueta entidade são termos que deverão ser encontrados nas informações dos documentos no instante em que houver uma comparação entre os termos digitados e os termos presentes nos metadados. Pode-se aplicar a mesma importância de uma palavra-chave nos termos que são identificados como entidades. É exatamente neste momento em que a presença dos metadados permite o enriquecimento do diálogo com o usuário, fornecendo termos de qualidade provenientes do momento da inclusão do documento no repositório por um usuário qualificado que tem o conhecimento a respeito do conteúdo de tais documentos. Se uma inteligência artificial processa os termos, esta não é capaz de separar e atribuir com a devida qualidade os termos escolhidos, pois no caso deste exemplo e aplicado neste trabalho, no momento da submissão do trabalho é o próprio autor que escolhe os que serão introduzidos na plataforma, e quem mais poderia ter um maior conhecimento a respeito de tal documento do que o próprio autor. Mesmo sistemas de aprendizagem de máquina podem fazer um ótimo

trabalho na recuperação de informações, mas o processo cognitivo da escolha dos termos que representam um documento não pode ser simulado, determinar que um termo é mais importante que outro e que deve vir em primeiro lugar, ou não, faz parte do aprendizado humano, porém não serão discutidos neste trabalho os processos envolvidos.

Quando o usuário requer, por exemplo, informações como um autor de trabalhos sobre tesauros, neste instante os termos **autor** e **tesauros** são os termos identificados como entidades e estes farão parte do processo de busca dos trabalhos. Formulando uma consulta ao banco com os parâmetros “**autor**” + “**tesauros**” muitos trabalhos serão recuperados e apresentados para a aplicação que processará os resultados, às vezes, termos compostos podem ser utilizados, como “**recuperação de informação**” e “**inteligência artificial**” e devem ser processados para que os termos não possam ser separados ou reduzidos para que a semântica seja mantida e não influencie no processo de recuperação dos documentos.

Ainda sobre os termos que foram extraídos, existe a situação em que no diálogo atual não existem mais termos fornecidos, mas estes devem estar presentes, dos diálogos anteriores, na forma de contextos. São casos em que mesmo na ausência do termo, sua presença é bastante importante e deve ser garantida que estará presente na construção dos parâmetros para a busca no repositório. O Quadro 13 mostra todos os termos que foram recuperados sob a etiqueta de contexto no diálogo.

Quadro 13 – Contextos

Contexto:	Significado
2021	Herda informações sobre ano da qual deseja o trabalho ou um trabalho sobre algo desse ano.
filmes	Herda informações sobre assunto da qual deseja o trabalho ou um trabalho sobre algo desse assunto.
inteligência artificial	Herda informações sobre assunto da qual deseja o trabalho ou um trabalho sobre algo desse assunto.
natureza	Herda informações sobre assunto da qual deseja o trabalho ou um trabalho sobre algo desse assunto.
trabalho	Herda informações sobre o trabalho citado.
trabalhos	Herda informações sobre os trabalhos citados.

Fonte: elaborado pelo autor

Como pode-se verificar, a maioria dos termos rotulados como **contexto**, em um diálogo anterior foram **entidades**, sendo uma forma de garantir que o assunto possa fluir durante diversos diálogos sem perder de foco quais termos estão em evidência. Um termo de **contexto** deve ser manipulado com mais de uma variável na construção das expressões de busca, com

garantias que o seu estado possa transitar entre um diálogo e outro, até que um novo termo seja utilizado para referenciar a mesma variável.

Uma informação muito importante a ser enfatizada, neste momento, é a necessidade da criação de coeficientes de ponderação para cada domínio do texto, estes que podem ser: título, palavra-chave, resumo, autores e conteúdo. Isso é relevante devido ao fato que em diferentes campos da literatura científica existem graus diferentes de generalização. O título pode indicar o assunto e o objetivo da pesquisa. As palavras-chave podem utilizar de termos para sintetizar a implementação técnica envolvida. O resumo tem por razão descrever brevemente o objetivo, o método e o resultado da pesquisa. O conteúdo pode ser descrito como a expansão de operações específicas da literatura. Os autores possuem um grau diferente de valor devido ao fato de não utilizarem termos de nenhuma literatura, mas seu grau de importância é elevado, pois denota a origem e a ligação entre os trabalhos. Não é possível realizar o mesmo processo para a correspondência entre todas as partes enumeradas e as palavras da pesquisa (RUI *et al.*, 2017). Diferentes coeficientes de peso precisam ser desenvolvidos para serem introduzidos nessa equação para que a recuperação das informações seja de maneira mais eficiente. Até mesmo a ordem das palavras cadastradas nos campos das palavras-chave precisa ser avaliada, pois para o autor, a ordem é importante para expressar o assunto e o grau de comprometimento de cada termo no desenvolvimento do texto. Ao se tratar dos autores, a ordem poderia indicar o grau de participação de cada nome e a quantidade de contribuição no trabalho. Pode ser a diferença entre um autor de um coautor e essa informação pode ser utilizada para responder algum questionamento de um usuário do agente ao perguntar mencionando um nome para saber se ele era o autor principal ou havia participado de algum trabalho como coautor.

Ao determinar valores de pesos diferentes para termos como título, palavra-chave, autores, ano, resumo, nome e código de grupo de trabalho, a taxa de sucesso na encontrabilidade de um documento seja considerada maior do que se os termos utilizados na pesquisa fossem comparados com os termos do conteúdo do documento. Visto que a quantidade de termos é muito maior e praticamente qualquer documento poderia ter relação com os termos utilizados mesmo com uma relevância muito baixa fazendo com que tais documentos fossem recuperados e utilizados como resultados para apresentação ao usuário do agente.

Quanto maior a organização na criação dos termos dos metadados, e ainda mais se sua origem for ligada a vocabulários controlados, mais resultados positivos com menos interações são possíveis, o que é ideal para utilização em sistemas conversacionais.

Ao tratar de interfaces, a partir dos textos criados para esta simulação, para demonstrar que um mesmo repositório utilizado para este projeto, o ENANCIB WEB, que se encontra disponível na Web, permitiu desenvolver pelo menos 3 formas diferentes de acesso por meio de texto. Ao configurar o agente para ser utilizado por meio do Dialogflow, foram escolhidas as seguintes integrações:

- Página dedicada no sítio ENANCIB WEB
- BOT nas páginas do sítio ENANCIB WEB
- Telegram *Desktop* (utilizado em Computador de Mesa com Windows 11)
- Telegram Celular (utilizado em celular com Android)

Todos os diálogos entre o usuário e o agente, baseados nos textos apresentados acima, foram utilizados para simular uma conversa com o agente por meio das respectivas interfaces, foram apresentados os questionamentos ao *chatbot* e este respondeu de forma esperada de acordo com as informações provenientes do banco de dados do repositório, e todas as telas contendo estes diálogos podem ser visualizados por meio dos anexos deste trabalho (I, II e III).

O Anexo I permite visualizar todas as interações por meio da página desenvolvida e hospedada em <http://ricdon.com.br/enancib/chatbot>, nela também é possível verificar no canto inferior direito um pequeno logotipo da cor azul e laranja que permite acessar uma pequena janela do agente que possui os mesmos parâmetros da página supracitada, a diferença é que este ícone, e suas funcionalidades, podem ser introduzidas em todas as páginas do sítio ENANCIB WEB e permitir que a qualquer momento o agente possa ser invocado.

O Anexo II permite verificar as interações com o agente por meio do *software* Telegram instalado em um computador que utiliza Windows 11, onde permite a interação e pode responder às perguntas da mesma forma que o formato anterior, porém todo o tráfego é feito de forma criptografada e por meio da infraestrutura do mantenedor do aplicativo de mensagens.

Através do Anexo III, pode-se verificar a utilização do aplicativo de mensagens Telegram por meio de um celular que utiliza o sistema operacional Android. Da mesma maneira que é possível interagir com outras pessoas, o agente está disponível para ser indagado a qualquer momento e de qualquer lugar. Em casos em que utiliza um aplicativo de terceiros, como de mensagens ou redes sociais, as conversas precisam passar por todas as etapas, como demonstrado anteriormente, inicialmente processada pela plataforma do Telegram, em seguida é recebida pelo módulo de integração do Dialogflow onde os termos são extraídos e enviados ao módulo de Ações, que pode ser interno ou externo à ferramenta. Na próxima etapa os dados

são recuperados do banco de dados e devolvidos ao módulo de diálogo para que possam ser reconstruídos e enviados por meio do módulo de integração ao Telegram e este apresente o resultado na forma de respostas ao usuário.

Os modelos principais em que se dividem as taxonomias dos *chatbots* são os baseados em regras e os generativos. Os sistemas baseados em regras possuem toda a inteligência do domínio codificado em regras, onde a partir da comparação de palavras, existem perguntas possíveis e as respectivas respostas. Já um sistema generativo exige alto grau de inteligência artificial e aprendizagem de máquina para alcançar todos os termos que são enviados pelo usuário e processá-los, assim como todos os documentos que compõem a base de informações do sistema para encontrar a resposta mais provável para a pergunta do usuário, mas nesse caso o sistema acaba ficando bastante complexo e demanda muita tecnologia e profissionais especializados.

O modelo proposto é um híbrido de um sistema baseado em regras, mas que consegue fornecer algumas funções que seriam exclusivas dos sistemas generativos, porém sem utilizar altos níveis de processamento de linguagem natural, inteligência artificial e aprendizado de máquina no processamento dos resultados, porque por meio dos metadados é possível encontrar as relações entre os termos de um domínio específico e enriquecer cada resultado durante as interações com usuário. Qualquer vocabulário controlado que for necessário para o funcionamento deste agente já existe e é encontrado na forma dos metadados, e ainda melhor, se estes forem baseados em um tesauro ou ontologia de domínio do *corpus*. Nesse caso, o processo de levantamento das entidades e das intenções podem se beneficiar de maneira bastante positiva devido a essa padronização.

5.2 Avaliação do Modelo e Interface

Um sistema como este precisa ser avaliado constantemente a respeito da qualidade dos resultados, de forma a diminuir ao menor número possível de resultados entregues aos usuários com respostas que não atendam às expectativas destes. Ainda não existem padrões ou boas práticas para avaliação de sistemas de conversação, o que é utilizado neste momento são as ferramentas que as próprias empresas desenvolvedoras fornecem para avaliar os resultados baseados nas perguntas, permitindo aos analistas verificar o grau de acerto e fazer todas as correções necessárias nas configurações do agente e melhorando as interações conforme o sistema é utilizado. No caso do Dialogflow, são os módulos de *Analytics* e *History* que são

disponibilizados com relatórios detalhados para esse propósito. A partir de 2017, houve um grande avanço no desempenho de sistemas de processamento de linguagem natural, o que permitiu que muitos sistemas de criação de agentes pudessem usufruir dessa tecnologia e avançar nessa área. O fato de ter acesso aos dados digitados pelos usuários tornam esse processo uma fonte formidável de informações, mas por serem descritas em linguagem natural nos diálogos, ainda é necessário sistemas avançados de aprendizagem de máquina e inteligência artificial para processar grandes volumes de dados e extrair informações que possam contribuir de maneira positiva.

Nessa busca de encontrar ferramentas para serem utilizadas no desenvolvimento de agentes inteligentes, muito se fala a respeito das 10 heurísticas de usabilidade desenvolvidas por Jakob Nielsen (GOMES, 2019; SCOTT, 2016; NIELSEN, 1994; MAMGAIN, 2022; SEN, 2022; SEO, 2017).

De uma forma resumida, Gomes (2019) descreve tais heurísticas como:

1. **Visibilidade do status do sistema** - O usuário deverá ser informado do que está acontecendo.
2. **Equivalência entre o sistema e o mundo real** - O sistema deve falar a linguagem do usuário, com palavras, ícones e conceitos familiares ao usuário, ao invés de termos técnicos que façam sentido apenas para desenvolvedores.
3. **Controle e liberdade do usuário** - Usuários frequentemente escolhem uma opção acidentalmente e devem ter a liberdade de voltar, ignorar ou avançar em qualquer etapa do sistema.
4. **Consistência e Padrões** - Os usuários não devem se perguntar se palavras, situações ou ações diferentes significam a mesma coisa.
5. **Prevenção de erros** - “Ainda melhor que uma boa mensagem de erro é um design cuidadoso que possa prevenir esses erros”. Elimine as condições propensas a erros ou verifique-as e apresente aos usuários uma opção de confirmação antes de confirmarem a ação.
6. **Reconhecimento ao invés de memorização (ou relembração)** - Minimize a carga de memória do usuário, tornando objetos, ações e opções visíveis. O usuário não deve ter que lembrar informações de uma parte do diálogo para outra.
7. **Flexibilidade e eficiência de uso** - O sistema deve ser ágil para os usuários experientes e flexível para os usuários leigos. E ainda, permitir automatizar tarefas frequentes.

8. **Estética e Design minimalista** - Os diálogos não devem conter informações irrelevantes ou raramente necessárias. Cada unidade extra de informação em um diálogo compete com as unidades relevantes de informação e diminui sua visibilidade relativa.
9. **Auxiliar usuários a reconhecer, diagnosticar e se recuperar de erros** - Mensagens de erro devem ser expressas em linguagem simples (sem códigos), indicar precisamente o problema e sugerir de forma construtiva uma solução.
10. **Ajuda e documentação** - Mesmo que seja melhor se o sistema puder ser usado sem documentação, pode ser necessário fornecer ajuda e documentação. Qualquer informação desse tipo deve ser fácil de pesquisar, focada na tarefa do usuário, listar as etapas concretas a serem executadas e não ser muito grande.

Ainda complementa que apesar das Heurísticas de Nielsen terem sido delineadas para a avaliação de usabilidade de páginas Web, na década de 90, tem sido bem empregada independentemente do tipo de interface, devido ao fato de terem sido utilizadas no desenvolvimento das interfaces mobile, e neste momento tem sido utilizada com muito sucesso na avaliação de interfaces conversacionais.

No Quadro 14, se discute como as Heurísticas de Nielsen podem orientar a criação de métricas para o monitoramento do *chatbot*, ou agente, fornecendo recursos adicionais para o desenvolvimento destes e aprimoramento da utilização na busca da eficiência.

Quadro 14 – Proposta de Avaliação por Nielsen

Heurísticas de Nielsen		Modelo da Pesquisa
1	Visibilidade do status do sistema	O agente deve manter o usuário informado sobre o que está acontecendo e permitir que este pergunte o status do sistema a qualquer momento e receba uma resposta adequada no tempo adequado.
2	Equivalência entre o sistema e o mundo real	Neste caso o sistema deve falar uma linguagem similar, de forma que toda a informação que ele receber do sistema pareça que é natural e familiar.
3	Controle e liberdade do usuário	O usuário pode ter a liberdade de escrever o que deseja, mas em casos em que por engano algo é digitado, deve haver sistemas de controle para que o usuário possa voltar ao diálogo sem perder a linha da comunicação.
4	Consistência e Padrões	É necessário seguir um padrão para manter a consistência da linguagem, se for por linguagem natural, toda a comunicação deve ser dessa forma, se houver a utilização de ícones, que estes sejam padronizados para não confundir o usuário.
5	Prevenção de erros	Para evitar que o usuário cometa erros, deve pedir confirmações ou apresentar opções para refazer uma tarefa, em todos os momentos críticos do diálogo.
6	Reconhecimento ao invés de memorização	Para não exigir que o usuário se lembre de todas as informações, ou partes dela; é admissível a utilização de ícones, menus, botões, entre outros recursos, de forma a deixar de maneira visível as informações que o usuário pode precisar.
7	Flexibilidade e eficiência de uso	O sistema deve possuir a habilidade de ser flexível ao ponto de conseguir atender desde o usuário mais experiente até o mais leigo. Ou seja, de

		usuários que digitam frases bastante elaboradas contendo muitos termos até outros usuários que expressam o seu desejo por apenas poucas palavras, mas que para o sistema, ambos serão entendidos.
8	Estética e Design minimalista	Ao garantir que os diálogos contenham apenas as informações que sejam relevantes, além de conteúdo minimalista, permite ao usuário uma melhor compreensão dos dados que lhes são apresentados.
9	Auxiliar usuários a reconhecer, diagnosticar e se recuperar de erros	Em casos em que o usuário tenha inserido alguma informação incorreta, o sistema deve informá-lo do que está acontecendo e permitir que o usuário, ao identificar o seu erro, possa corrigi-lo, evitando ao máximo durante esse processo a utilização de códigos que possam ser impeditivos para utilização do usuário.
10	Ajuda e documentação	Toda documentação que houver a respeito do agente deve estar disponível para utilização pelo próprio agente. Em momentos que o usuário precisar de ajuda, é importante listar opções que podem auxiliá-lo a entender o que está acontecendo e decidir os próximos passos.

Fonte: elaborado pelo autor

Com o propósito de definir algumas diretrizes sobre a realização de um processo básico de avaliação para um *chatbot*, Sen (2022) elenca algumas razões:

- Para determinar se o sistema funciona como esperado;
- Para determinar se o sistema atende as necessidades do usuário, entende suas declarações e o ajuda a resolver suas consultas, por meio de uma experiência satisfatória;
- Para estabelecer se os objetivos propostos foram cumpridos.

Conclui-se que a avaliação de sistemas de diálogo é uma tarefa importante e complexa, com muitas questões que devem ser consideradas. Com o propósito de avaliar a qualidade da conversa e a satisfação dos usuários, algumas dessas métricas precisam ser definidas e aplicadas durante o processo de desenvolvimento para que depois possam ser avaliadas, corrigidas ou melhoradas. Neste momento, não faz parte desta pesquisa a aplicação de métricas de avaliação sobre o modelo, mas é importante mencionar que existem trabalhos a respeito desse processo e muitos avanços deverão acontecer nos próximos anos, acompanhando o desenvolvimento das ferramentas de criação de agentes conversacionais.

6.

CONSIDERAÇÕES FINAIS

Durante os anos de desenvolvimento desta pesquisa, foi interessante observar que no início se falava, tanto em artigos científicos, quanto em revistas e *sites*, que o futuro seria baseado em *chatbots* nos processos de comunicação com usuários ou clientes, havia uma grande expectativa a respeito de vários produtos que estavam sendo lançados e que começaram a ser utilizados por pequenas e grandes empresas, e até mesmo pessoas físicas, e com o passar do tempo, muitas empresas desenvolvedoras de ferramentas que são utilizadas para a criação dos agentes foram absorvidas por outras empresas, ou mesmo desapareceram. Parece até que estava se formando uma bolha dos serviços de agentes conversacionais, a exemplo de outras bolhas que já aconteceram no passado. Mas durante os anos de 2020 e 2021, o foco mudou e todos entenderam melhor o que era possível, e como, ser feito por meio de agentes disponíveis na Web, em redes sociais ou aplicativos de mensagens.

Neste momento, em 2022, as ferramentas que estão disponíveis para os desenvolvedores de agentes estão mais maduras e diversos conceitos foram aperfeiçoados, permitindo a construção de agentes, até mesmo mais simples e utilizando interfaces por meio de praticamente nenhuma linha de código, apenas com a utilização de blocos, nos quais já é possível criar um sistema que pode interagir com os usuários e responder questões mais simples. Algumas ferramentas comerciais desse tipo começaram a receber um novo olhar e novos investimentos, como o Google Duplex, que permite a criação de agentes que podem interagir com outros humanos por meio de ligações telefônicas simulando um ser humano, criado em maio de 2018.

Foi anunciado em 2022 no evento anual do Google, como um produto que receberá maior investimento por parte de seu mantenedor e possibilitando que a ferramenta faça parte de outros produtos, a disponibilização para que os desenvolvedores externos comecem utilizá-la na construção de agentes conversacionais mais avançados e a liberação para começar a ser utilizada no Brasil, mesmo de forma limitada.

Ao inovar a forma de utilizar metadados em processos de recuperação de informação, utilizando-os nos processos comunicacionais entre um usuário e o sistema, considera-se ter contribuído para a área da Ciência da Informação, principalmente para a Linha de Informação e Tecnologia da UNESP de Marília.

Esta pesquisa foi motivada pela observação da necessidade de se utilizar as novas tecnologias em interface usadas em sistemas de conversação, em redes sociais e aplicativos de mensagens, para acesso a informações disponíveis em repositórios na Web e que possuem metadados cadastrados, com o intuito de resolver alguns problemas de comunicação desse processo. Importante relembrar a questão a ser respondida por este trabalho: *"Como um chatbot pode ser utilizado para recuperar informações de um repositório Web, de forma que os seus metadados possam ser utilizados para resolver possíveis ambiguidades e imprecisões durante o processo de comunicação entre o usuário e o sistema?"*

Por meio de hipóteses como, se metadados poderiam ser utilizados na construção de fluxos conversacionais, se o relacionamento entre os termos descritos como metadados poderia melhorar a relevância de um diálogo, se seria possível utilizar uma estrutura de um portal existente e disponível na Web para fornecer tais metadados, e se durante um diálogo, os metadados poderiam fornecer novos termos, ou fazer ligações entre eles, e comunicar ao usuário para facilitar a recuperação de informações. Para corroborar tudo isso, esta pesquisa construiu a seguinte tese que “considerando a recuperação de informação como um processo comunicacional e dialógico, problemas inerentes à comunicação como subjetividade, imprecisão e ambiguidade podem ser minimizados quando os metadados que descrevem e representam os documentos em um repositório podem ser utilizados no desenvolvimento de agentes conversacionais”.

Este trabalho propõe a criação de um modelo conceitual de recuperação de informação por meio de um *chatbot*, a partir da qual, no momento da criação dos diálogos na comunicação com um usuário, os metadados dos documentos do repositório podem contribuir para enriquecer os termos para a construção da expressão de busca. Situando-se na convergência entre as áreas da Ciência da Informação, dos agentes conversacionais, dos sistemas de recuperação de informação e dos metadados.

Como objetivo geral norteando esta pesquisa no intuito de observar o efeito de um agente conversacional, como um *chatbot*, obtendo acesso aos metadados durante o processo de construção dos diálogos com o usuário, tendo este objetivo sendo cumprido com a criação do

modelo e o seu teste por meio da construção de um agente disponível por meio de três interfaces diferentes, testado ao ponto de comprovar que tal acesso aos metadados pode sim influenciar a qualidade dos resultados, com menor número de interações e menor utilização de tecnologias complexas de análise de dados.

Com relação aos objetivos específicos, primeiramente o objetivo era identificar na literatura oriunda da área da Ciência da Informação o atual cenário da pesquisa em interfaces e modelos de agentes conversacionais e suas aplicações, este objetivo foi cumprido com os capítulos que descrevem toda a teoria a respeito de tais informações, elencando os principais autores, trabalhos e tecnologias disponíveis e utilizadas para a construção e disponibilização de produtos baseados nessas tecnologias.

Em seguida, sobre verificar a aplicabilidade de interfaces em linguagem natural em sistemas de recuperação de informação, entre outros, foi cumprido com a pesquisa e escrita do capítulo sobre a recuperação de informação e os seus modelos clássicos.

Como terceiro objetivo de analisar as etapas de pré-processamento, processamento e pós-processamento do modelo de comunicação por meio de agentes conversacionais, o cumprimento se deu pela escrita do texto do capítulo quinto, pois para o desenvolvimento do modelo teórico da proposta, era necessário o entendimento de todas as etapas relativas ao processo de diálogo, desde o momento em que o usuário digita uma pergunta, até a consulta ao banco e a devolutiva em forma de linguagem natural da resposta contendo as informações que o usuário busca.

O quarto objetivo era verificar se o modelo poderia ter sucesso ao ser aplicado a um repositório com documentos e metadados cadastrados, cumprido ao ser desenvolvido um modelo de *chatbot* e aplicado à página na Web da Biblioteca Digital Brasileira de Teses e Dissertações (BDTD), permitindo resultados interessantes durante o diálogo.

Por último, como quinto objetivo a elaboração de um protótipo de agente conversacional com uma demonstração da viabilidade, usabilidade e desempenho desta proposta, foi cumprido com o desenvolvimento do portal ENANCIB WEB, que encontra-se finalizado e permite a busca por meio de termos em todos os trabalhos apresentados nos eventos desde a sua primeira edição, com os metadados que foram atribuídos aos documentos foi possível utilizá-los no protótipo AGENTE ENANCIB que pode interagir com os usuários por meio de uma página Web, pelo Telegram em computadores PC e no aplicativo Telegram dos celulares.

A respeito da contribuição social, nosso trabalho pode contribuir ao demonstrar que um sistema baseado no modelo proposto pode ser mais simples de desenvolver e mais barato de manter devido ao fato de não precisar de módulos avançados de processamento de inteligência, e assim permitir que mais repositórios de documentos possam ficar disponíveis por meio de ferramentas já conhecidas, como mensageiros instantâneos e redes sociais, e possibilitar a um público bastante eclético dessas redes, incluindo os com necessidades especiais, como os portadores de deficiência visual que podem usar a voz, além de utilizar a ferramenta para recuperar informações sem depender de uma curva de aprendizagem, que poderia impedir algumas pessoas de ter sucesso na sua busca.

Embora a maioria dos resultados serem positivos, foram identificadas algumas dificuldades que necessitam de maior dedicação para serem completamente sanadas. Ao desenvolver o modelo e trabalhar na construção do algoritmo do protótipo, o ranqueamento dos termos oriundos como resposta da consulta ao banco dependem de um sistema de ranqueamento eficiente e complexo, necessitando ter acesso aos termos das entidades, intenções e contextos, mas como a codificação pode ser feita por meio de uma API, esta tem dificuldades de acesso aos termos da ferramenta Dialogflow por motivo das transações serem *StateLess* (Sem Estado), uma funcionalidade que obriga que todas as transações de dados devem ser completas e com todos os dados, porque o sistema não armazena nenhuma informação ou parâmetro entre uma requisição e outra, é uma das restrições da tecnologia REST, considerada a melhor maneira de desenvolver esta integração.

Se a codificação estiver dentro da ferramenta Dialogflow, ocorre o mesmo, dificuldade de acessar aos termos que estão hospedados na API externa. Para demonstração do modelo, esta transação teve que ser feita manualmente, e este imbróglio deve ser concluído posteriormente a este trabalho, com a sugestão de não utilização de ferramentas como o Dialogflow, mas com o desenvolvimento completo do *chatbot* e hospedagem em servidores de propriedade do desenvolvedor, porém com a utilização de módulos de terceiros para fazer algumas etapas do processo de comunicação.

Como proposta de trabalhos futuros nesta solução, a conclusão de todos os módulos do processo para utilização de maneira autônoma, e teste com muitos usuários para avaliação do processo. O protótipo deve ser considerado como uma prova de conceito o qual precisa de desenvolvimento para se tornar um produto finalizado, o que deverá acontecer nos próximos anos com o prosseguimento desta pesquisa por parte do autor.

Como trabalhos futuros relacionados, poderão ser conduzidos estudos a respeito da criação de coeficientes de ponderação para cada domínio do texto, como título, palavra-chave, resumo, autores e conteúdo. Ao se tratar das pesquisas bibliográficas para este trabalho, muito pouco foi encontrado sobre estes assuntos e foi determinado que para alguns tipos de tarefas, é muito importante a classificação correta dos termos utilizados para a qualidade dos resultados.

O principal diferencial desta proposta é ser um projeto aplicado, no qual foram utilizados os preceitos oriundos da Ciência da Informação, das disciplinas da pós-graduação e das reuniões com o orientador, na tentativa da resolução de uma dificuldade do mundo real, que é a utilização de novas tecnologias em interfaces inovadoras em processos de recuperação de informação, na qual a presença de metadados pode ser um diferencial para o incremento da qualidade do resultado e diminuição dos custos em tecnologia. E por fim, comprova-se a tese de que ao submeter um *chatbot* a um ambiente com acesso a metadados é possível resolver a subjetividade pelo diálogo na recuperação de informação.

REFERÊNCIAS

ABDUL-KADER, Sameera A.; WOODS, J. C. Survey on chatbot design techniques in speech conversation systems. **International Journal of Advanced Computer Science and Applications**, v. 6, n. 7, 2015.

AHMAD, Amin. **The High Cost of Keyword Search**. 2021. Disponível em: <<https://blog.zir-ai.com/the-high-cost-of-keyword-search>>. Acesso em 19 set. 2021.

ALVARES, Lillian; ARAÚJO JÚNIOR, Rogério Henrique de. Marcos históricos da ciência da informação: breve cronologia dos pioneiros, das obras clássicas e dos eventos fundamentais. **Transinformação**, v. 22, p. 195-205, 2010.

ALVES, Rachel Cristina Vesu. **Metadados como elementos do processo de catalogação**. 2010. 132 f. Tese (Doutorado em Ciência da Informação) – Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília, 2010. Disponível em: <http://www.marilia.unesp.br/Home/Pos-Graduacao/CienciadaInformacao/Dissertacoes/alves_rcv_do_mar.pdf>. Acesso em: 17 dez. 2021.

ALVES, Rachel Cristina Vesu; SANTOS, Plácida Leopoldina Ventura Amorim da Costa. **Metadados no Domínio Bibliográfico**. Rio de Janeiro: Intertexto, 2013.

ARAMPATZIS, et al. Linguistically-motivated Information Retrieval. **Encyclopedia of Library and Information Science**, v. 69, p. 201-222, 2000.

AZZOPARDI, Leif et al. Conceptualizing agent-human interactions during the conversational search process. In: **The Second International Workshop on Conversational Approaches to Information Retrieval**. 2018.

BACA, Murtha (ed.). **Introduction to Metadata**. 2. ed. Los Angeles: Getty Research Institute, 2008.

BAEZA-YATES, Ricardo; RIBEIRO-NETO, Berthier. **Modern Information Retrieval**. 2ª ed. Addison-Wesley, 2011.

BARTL, Alexander; SPANAKIS, Gerasimos. A retrieval-based dialogue system utilizing utterance and context embeddings. In: **2017 16th IEEE International Conference on Machine Learning and Applications (ICMLA)**. IEEE, 2017. p. 1120-1125.

BEARDON, Colin; LUMSDEN, David; HOLMES, Geoff. **Natural language and computational linguistics**. Melksham-Wiltshire, England: Ellis Horwood, 1991.

BERNERS-LEE, Tim. **Tim Berners-Lee: i invented the web. here are three things we need to change to save it. I invented the web. Here are three things we need to change to save it.** 2017. Disponível em: <https://www.theguardian.com/technology/2017/mar/11/tim-berners-lee-web-inventor-save-internet>. Acesso em: 28 dez. 2021.

BERNERS-LEE, Tim. **Metadata Architecture.** 1997. Disponível em: <https://www.w3.org/DesignIssues/Metadata.html>. Acesso em: 20 dez. 2021.

BORKO, Harold. Information science: what is it?. **American documentation**, v. 19, n. 1, p. 3-5, 1968.

BRIET, Suzanne. **What is documentation?** Trad. de Ronald E. Day; Laurent Martinet; Hermina G.B. Anghelescu. Lanham, MD: Scarecrow Press, 2006.

BUCKLAND, Michael Keeble. **Information and Information Systems.** New York: Greenwood, 1991.

BUSH, Vannevar et al. As we may think. **The atlantic monthly**, v. 176, n. 1, p. 101-108, 1945.

CAHN, Jack. CHATBOT: Architecture, design, & development. University of Pennsylvania School of Engineering and Applied Science Department of Computer and Information Science, 2017.

DE BARROS CAMPOS, Luiz Fernando. Metadados digitais: revisão bibliográfica da evolução e tendências por meio de categorias funcionais. **Encontros Bibli: revista eletrônica de biblioteconomia e ciência da informação**, n. 23, p. 16-46, 2007.

CARVALHO, Maria Margarida Melo de O. problema da subjetividade na indexação. **Cadernos BAD**, v. 1, p. 9-40, 1995.

CARVALHO, Ricardo César de. **Aplicação de técnicas de mineração de texto na recuperação de informação clínica em prontuário eletrônico do paciente.** 2017. 201 f. Dissertação (Mestrado em Ciência da Informação) – Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília, 2017.

CASAROTTO, Camila. **Entenda o que são Rich Snippets e sua importância para SEO.** 2017. Disponível em: <<https://rockcontent.com/br/blog/rich-snippets>>. Acesso em: 20 jan. 2022.

CETIC.BR. **Resumo Executivo TIC Domicílios 2019.** 2020a. Disponível em <https://www.cetic.br/media/docs/publicacoes/2/20201123115919/resumo_executivo_tic_dom_2019.pdf> Acesso em: 27 out. 2021.

CETIC.BR. **TIC DOMICÍLIOS 2019.** 2020b. Disponível em <https://www.cetic.br/media/analises/tic_domicilios_2019_coletiva_imprensa.pdf> Acesso em: 27 out. 2021.

CETIC.BR. **TIC DOMICÍLIOS 2021.** 2022. Disponível em <https://www.cetic.br/media/analises/tic_domicilios_2021_coletiva_imprensa.pdf> Acesso em: 27 ago. 2022.

CIBANGU, Sylvain K. Communication science and information science: Convergences and divergences. **Emporia State Research Studies**, v. 50, n. 1, p. 22-35, 2015.

CIO. **Em 2018, chatbots vão ampliar conversa entre marcas e consumidor**. 2018. Disponível em: <<https://cio.com.br/tendencias/em-2018-chatbots-vaio-ampliar-a-conversa-entre-marcas-e-consumidor/>>. Acesso em: 08 jan. 2022.

CORREIA, Marcelo Francisco de Barros. **RECUPERAÇÃO DE INFORMAÇÃO**. Nairobi: African Virtual University, 2017. 108p.

CRESTANI, Fabio; PASI, Gabriella. Handling Vagueness, Subjectivity, and Imprecision in Information Access: An Introduction to the Special Issue. **Information Processing & Management**, v. 39, n. 2, p. 161-165, 2003.

DCMI. **Dublin Core Metadata Initiative**. 2021. Disponível em: <<https://www.dublincore.org/>>. Acesso em: 20 dez. 2021.

DE BARROS CAMPOS, Luiz Fernando. Metadados digitais: revisão bibliográfica da evolução e tendências por meio de categorias funcionais. **Encontros Bibli: revista eletrônica de biblioteconomia e ciência da informação**, n. 23, p. 16-46, 2007.

DENNING, Peter J.. et al. Computing as a discipline. **Computer**, v. 22, n. 2, p. 63-70, 1989.

DIAS, Danielle. Dialogflow: plataforma para desenvolvimento de chatbots. Plataforma para desenvolvimento de ChatBots. 2020. Disponível em: <https://embarcados.com.br/dialogflow-plataforma-para-desenvolvimento-de-chatbots/>. Acesso em: 20 jan. 2022.

DIAS, Paulo. **A abordagem da comunicação multidimensional na concepção e desenvolvimento de interfaces hipermedia**. 1994. Disponível em: <http://www.ufrgs.br/niee/eventos/RIBIE/1994/II_30_40.htm>. Acesso em 19 set. 2019.

FALOUTSOS, Christos; OARD, Douglas W. **A survey of information retrieval and filtering methods**. Technical Report CS-TR-3514. Department of Computer Science, University of Maryland, 1998.

FERNEDA, Edberto; DIAS, Guilherme Ataíde. A Lógica Fuzzy aplicada à recuperação de informação. **Revista InterScientia**, v. 1, n. 1, p. 51-65, 2013.

FERNEDA, Edberto. **Ontologia como recurso de padronização terminológica em um Sistema de Recuperação de Informação**. 2013. 97 f. Tese (Pós-Doutorado) - Curso de Ciência da Informação, Universidade Federal da Paraíba, João Pessoa, 2013.

FERNEDA, Edberto. **Introdução aos Modelos Computacionais de Recuperação de Informação**. Rio de Janeiro: Ciência Moderna, 2012.

FERNEDA, Edberto; DIAS, Guilherme Ataíde. OntoSmart: proposta de um modelo de recuperação de informação baseado em ontologia. In: **II Congreso ISKO España-Portugal. Organización del conocimiento: sistemas de información abiertos**. Universidad de Murcia, 2015. p. 105-115.

FURRIE, Betty. **Understanding MARC Bibliographic: Machine-Readable Cataloging**. 2009. Disponível em: <<https://www.loc.gov/marc/umb/>>. Acesso em: 20 dez. 2021.

GIL, Antonio Carlos. **Como elaborar projetos de pesquisa**. 4. ed. São Paulo: Atlas, 2002.

GOMES, Henrique. **Heurísticas de Nielsen para Bots**. 2019. Disponível em: <<https://brasil.uxdesign.cc/heur%C3%ADsticas-de-nielsen-para-bots-60363363d12>>. Acesso em 03 dez. 2020.

GOOGLE. **Conceitos básicos do Dialogflow ES**. 2022. Disponível em: <https://cloud.google.com/dialogflow/es/docs/basics>. Acesso em: 20 jan. 2022.

GOOGLE TRENDS. **Comparar**. 2021. Disponível em: <<https://trends.google.com.br/trends/explore?date=2004-01-01%202021-09-30&q=chatbot,chatterbot>>. Acesso em: 30 set. 2021.

IBM. **Computação cognitiva e tecnologias de fala são tema de evento realizado pelo CPqD em parceria com IBM**. 2018. Disponível em: <<https://www.ibm.com/blogs/robertoa/2017/06/computacao-cognitiva-e-tecnologias-de-fala-sao-tema-de-evento-realizado-pelo-cpqd-em-parceria-com-ibm/>>. Acesso em: 19 set. 2018.

INGWERSEN, Peter. **Information retrieval interaction**. London: Taylor Graham, 1992.

JACOB, Elin K.; SHAW, Debora. Sociocognitive Perspectives on Representation. **Annual Review of Information Science and Technology (ARIST)**, v. 33, p. 131-85, 1998.

JACQUEMIN, Christian; KLAVANS, Judith L.; TZOUKERMANN, Evelyne. Expansion of multi-word terms for indexing and retrieval using morphology and syntax. In: **35th Annual Meeting of the Association for Computational Linguistics and 8th Conference of the European Chapter of the Association for Computational Linguistics**. 1997. p. 24-31.

JOHNSTON, Bill; WEBBER, Sheila. As we may think: Information literacy as a discipline for the information age. **Research strategies**, v. 20, n. 3, p. 108-121, 2005.

KAUSHIK, Abhishek. Dialogue-Based Information Retrieval. In: **European Conference on Information Retrieval**. Springer, Cham, 2019. p. 364-368.

LE COADIC, Yves-François. **A Ciência da Informação**. 2.ed. Brasília: Briquet de Lemos, 2004.

LEVIATHAN, Yaniv; MATIAS, Yossi. Google Duplex: An AI System for Accomplishing Real-World Tasks Over the Phone. **Google AI Blog**, 8 mai. 2018. Disponível em: <<https://ai.googleblog.com/2018/05/duplex-ai-system-for-natural-conversation.html>>. Acesso em: 08 set. 2020.

LEWIS, David D.; JONES, Karen Spärck. Natural language processing for information retrieval. **Communications of the ACM**, v. 39, n. 1, p. 92-101, 1996.

LIMA, Gercina Ângela Borém. Interfaces entre a ciência da informação e a ciência cognitiva. **Ciência da Informação**, v. 32, p. 77-87, 2003.

LIU, Z. et al. Conversation in IR: its role and utility. In: **SIGIR Workshop on Conversational Approaches to IR**. 2017.

LOMMATZSCH, Andreas; KATINS, Jonas. An Information Retrieval-based Approach for Building Intuitive Chatbots for Large Knowledge Bases. In: **LWDA**. 2019. p. 343-352.

MAKHALOVA, Tatiana; ILVOVSKY, Dmitry; GALITSKY, Boris. Information Retrieval Chatbots Based on Conceptual Models. In: **International Conference on Conceptual Structures**. Springer, Cham, 2019. p. 230-238.

MAMGAIN, Devashish Datt. **12 Metrics For Chatbot Analytics You Must Track in 2022**. 2022. Disponível em: < <https://chatbotslife.com/12-metrics-for-chatbot-analytics-you-must-track-in-2022-4ec08ef10bdb>>. Acesso em 13 abr. 2022.

MARCONDES, Carlos Henrique. Representação e economia da informação. **Ciência da informação**, v. 30, p. 61-70, 2001.

MARCHIONINI, Gary. **Information seeking in electronic environments**. Cambridge university press, 1997.

MARON, Melvin Earl; KUHNS, John Larry. On relevance, probabilistic indexing and information retrieval. **Journal of the ACM (JACM)**, v. 7, n. 3, p. 216-244, 1960.

MARTINES, Alexandre ROBSON; OLIVEIRA, Lais Pereira de; SABBAG, Deise Maria Antonio. Subjetividade, conceito e representação da informação. In: **Estudos Avançados em Organização do Conhecimento. v. 5** - Organização do Conhecimento responsável: promovendo sociedades democráticas e inclusivas, p. 11-18, 2019.

MEADOW, Charles T. et al. **Text Information Retrieval System**. 3rded. London UK: Elsevier, 2007.

MEDIUM. **A Tech Stack de um Chatbot.**, 2020. Disponível em: <<https://medium.com/botsbrasil/a-tech-stack-de-um-chatbot-65a910217dd6>>. Acesso em: 29 ago. 2020.

MEY, Marc de. **The cognitive paradigm**: na integrated understanding of scientific development Chicago: University of Chicago, 1992.

MILSTED, Jessica, FELDMAN, Susan. **Metadata: cataloging by any other name**. ONLINE, jan. 1999. Disponível em: <<http://www.online.com/onlinemag/O11999/milstead1.html>>. Acesso em: 20 dez. 2018.

MOOERS, Calvin N. The theory of digital handling of non-numerical information and its implications to machine economics. Zator Company, 1950.

MOOERS, Calvin N. Zatocoding applied to mechanical organization of knowledge. **American documentation**, v. 2, n. 1, p. 20-32, 1951. Wiley-Blackwell. Disponível em: <<http://dx.doi.org/10.1002/asi.5090020107>>. Acesso em: 08 set. 2019.

NATIONAL SCIENCE FOUNDATION. Science Information specialists. **Proceedings of the Conference on Training Science Information specialists**, 1961, october, 1962, april, 12-13. Atlanta: Georgia Institute of Technology, 1961,1962. 139 p.

NICHOL, Alan. **The next generation of AI assistants in enterprise**. O'REILLY, 20 ago. 2018. Disponível em: <<https://www.oreilly.com/radar/the-next-generation-of-ai-assistants-in-enterprise/>>. Acesso em: 08 set. 2020.

NIELSEN, Jakob. **10 Usability Heuristics for User Interface Design**. 1994. Disponível em: <<https://www.nngroup.com/articles/ten-usability-heuristics/>>. Acesso em 03 dez. 2020.

NOVELLINO, Maria Salet Ferreira. Instrumentos e metodologias de representação da informação. **Informação & Informação**, v. 1, n. 2, p. 37-45, 1996.

ORACLE. **O que é um Assistente Digital?** 2020. Disponível em: <<https://www.oracle.com/br/solutions/chatbots/what-is-a-digital-assistant.html>>. Acesso em: 08 set. 2020.

ORENGO, Viviane Moreira; HUYCK, Christian R. A Stemming Algorithm for the Portuguese Language. In: **Proceedings of SPIRE2001 Symposium on String Processing and Information Retrieval**. 2001. p. 186-193.

POPPER, Karl. **Conjectures and refutations: The growth of scientific knowledge**. 4th rev ed. New York: Basic Books, 1972.

PÖTTKER, Luciana Maria Vieira. **Arquitetura para recuperação de objetos de aprendizagem** – uma abordagem baseada em agentes inteligentes e relevance feedback. 203 f. Tese apresentada ao Programa de Pós-Graduação em Ciência da Informação – Faculdade de Filosofia e Ciências da Universidade Estadual Paulista – UNESP, Marília, 2017.

RAJ, Sumit. **Construindo Chatbots com Python: Usando Natural Language Processing e Machine Learning**. [S.l.]: NOVATEC, 2019.

RIECKEN, Rinalda Francesca. Frame de temas potenciais de pesquisa em Ciência da Informação. **RDBCI: Revista Digital de Biblioteconomia e Ciência da Informação**, v. 4, n. 1, p. 43-63, 2006.

ROBERTSON, Stephen E.; JONES, K. Sparck. Relevance weighting of search terms. **Journal of the American Society for Information science**, v. 27, n. 3, p. 129-146, 1976.

ROBERTSON, Stephen E. Theories and models in information retrieval. **Journal of Documentation**, n. 33, p. 126-148, 1977.

ROTHERMEL, Alessandra; DE SOUZA DOMINGUES, Maria José Carvalho. MARIA: Um chatterbot desenvolvido para os estudantes da disciplina “Métodos e Técnicas de Pesquisa em Administração”. **SEGET-SIMPÓSIO DE EXCELÊNCIA EM GESTÃO E TECNOLOGIA**, Resende, 2007.

RUI, Zhang et al. Full-text Retrieval Model based on Term Frequency and Position Weighting. In: **2017 8th International Computer Systems and Education Management Conference (ICSEMC 2017)**. 2017. p. 31-38.

SACEANO, Daniel. **Chatterbots, Nanny-bots e outras criaturas**. 2000. Disponível em: <http://www.suite101.com/article.cfm/future_technology/6411> Acesso em: 09 mai 2018.

SALTON, Gerard; WONG, Anita; YANG, Chung-Shu. A vector space model for automatic indexing. **Communications of the ACM**, v. 18, n. 11, p. 613-620, 1975.

SANDERSON, Mark; CROFT, W. Bruce. The history of information retrieval research. **Proceedings of the IEEE**, v. 100, n. Special Centennial Issue, p. 1444-1451, 2012.

SARACEVIC, Tefko. A natureza interdisciplinar da ciência da informação. **Ciência da Informação**, v. 24, n. 1, 1995. p. 36-41.

SARACEVIC, Tefko. Ciência da informação: origem, evolução e relações. **Perspectivas em Ciência da Informação**, v. 1, n. 1, 1996.

SCHULTZ, Claire K. (ed.). **H.P. Luhn: Pioneer of information science - selected works**. New York: Spartan Books, 1968.

SCOTT, Kevin. **Usability Heuristics For Bots**. 2016. Disponível em: <<https://thekevinscott.com/usability-heuristics-for-bots/>>. Acesso em 10 jan. 2022.

SEN, Navaneeth. **Evaluating Chatbot Systems**. 2022. Disponível em: <<https://navaneethsen.medium.com/evaluating-chatbot-systems-81952e90401c>>. Acesso em 10 jan. 2022.

SEO, Eunji. **19 Best UX Practices for Building Chatbots**. 2017. Disponível em: <<https://chatbotsmagazine.com/19-best-practices-for-building-chatbots-3c46274501b2>>. Acesso em 15 mar. 2022.

SHEVAT, Amir. **Designing bots: Creating conversational experiences**. " O'Reilly Media, Inc.", 2017.

SMEATON, Alan F. Information retrieval: Still butting heads with natural language processing?. In: **International Summer School on Information Extraction**. Springer, Berlin, Heidelberg, 1997. p. 115-138.

SMIRAGLIA, Richard. **Metadata: a cataloger's primer**. Washington, DC: OUTLEDGE-USA, 2005.

STATCOUNTER. **Statcounter Global Stats**, 2022. Disponível em: <<https://gs.statcounter.com/platform-market-share/desktop-mobile-tablet/>> Acesso em: 02 jan 2022.

TURING, Alan M. **Computing machinery and intelligence**. Mind, v.59, n.236, p.433-460. 1950. Disponível em: <<https://academic.oup.com/mind/article/LIX/236/433/986238>>. Acesso em: 08 ago. 2020.

USP. **Projetos da USP sobre computação cognitiva recebem apoio da IBM**. 2018. Disponível em: <<http://jornal.usp.br/ciencias/ciencias-exatas-e-da-terra/projetos-da-usp-sobre-computacao-cognitiva-recebem-apoio-da-ibm/>>. Acesso em: 19 set. 2021.

VELLUCCI, Sherry L. Metadata. Annual Review of Information Science and Technology (ARIST), v. 33, p. 187-222, 1998.

VIEIRA, Simone Bastos. **La recuperación automática de información jurídica: metodología de análisis lógico-sintáctico para la lengua portuguesa**. 1994. 382 f. Tese (Doutorado em Ciência da Informação) - Universidad Complutense de Madrid, Madrid, 1994.

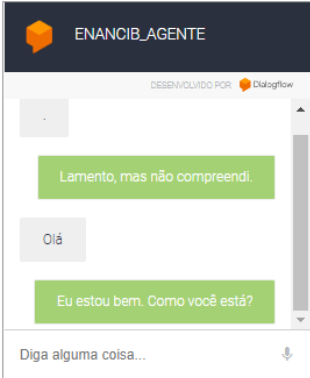
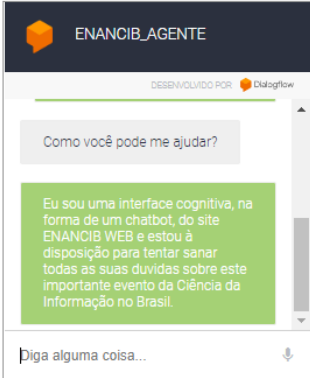
W3C. **Boas Práticas para Dados na Web**. 2017. Disponível em: <<https://w3c.br/traducoes/DWBP-pt-br>>. Acesso em: 20 dez. 2021.

WATSON: Transformando sua empresa com APIs cognitivas. New York: IBM Corporation, 2018.

WILSON, Thomas D. Human information behavior. **Informing science**, v. 3, p. 49-55, 2000.

ANEXO I

Neste Anexo são mostradas as interações efetuadas por meio de uma página disponibilizada no sítio ENANCIB WEB por meio do endereço <http://www.ricdon.com.br/enancib/chatbot> e respondendo às perguntas de um usuário, neste caso por texto. Além da existência de uma caixa de diálogo para a interação é possível fazer por meio de um pequeno ícone no canto direito inferior da página que pode ser adicionado à todas as páginas e permite fazer interações da mesma forma.

TEXTOS DOS DIÁLOGOS	IMAGENS DOS DIÁLOGOS VIA CHATBOT WEB
Olá Muito bem, obrigado.	 ENANCIB WEB via BOT 
Como você pode me ajudar? Eu sou uma interface cognitiva, na forma de um chatbot, do site ENANCIB WEB e estou à disposição para tentar sanar todas as suas dúvidas sobre este importante evento da Ciência da Informação no Brasil.	 ENANCIB WEB via BOT 

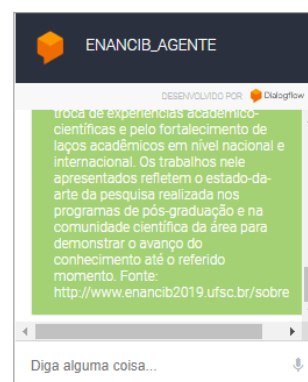
Você pode me falar mais sobre este evento?

O Encontro Nacional de Pesquisa em Ciência da Informação (ENANCIB) é o principal evento de pesquisa e de pós-graduação da área de Ciência da Informação do País e visa discutir e refletir a produção de conhecimento na área, de modo a estimular, por meio de amplo diálogo entre os pesquisadores que nela atuam, a realidade dos programas de pós-graduação. Consiste em um foro privilegiado para a reflexão e o compartilhamento das pesquisas científicas em Ciência da Informação e áreas de interface, congregando pesquisadores, docentes e pós-graduandos.

O evento é voltado à troca de experiências acadêmico-científicas e pelo fortalecimento de laços acadêmicos em nível nacional e internacional. Os trabalhos nele apresentados refletem o estado-da-arte da pesquisa realizada nos programas de pós-graduação e na comunidade científica da área para demonstrar o avanço do conhecimento até o referido momento. Fonte: <http://www.enancib2019.ufsc.br/sobre>



ENANCIB WEB via BOT



Quais os tipos de apresentações?

As modalidades são de Trabalho Completo e Resumo Expandido.



ENANCIB WEB via BOT

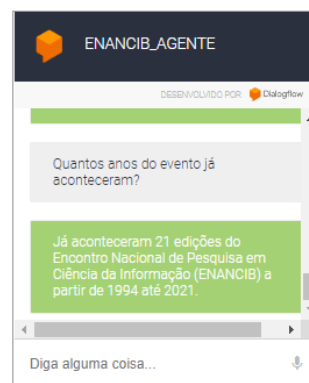


Quantos anos do evento já aconteceram?

Já aconteceram 21 edições do Encontro Nacional de Pesquisa em Ciência da Informação (ENANCIB) a partir de 1994 até 2021.



ENANCIB WEB via BOT

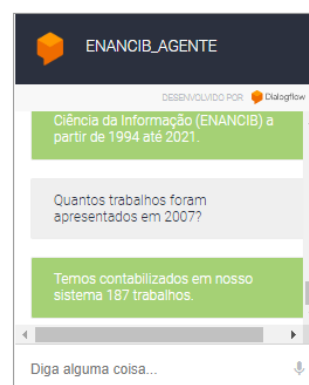


Quantos trabalhos foram apresentados em 2007?

Temos contabilizados em nosso sistema 187 trabalhos.



ENANCIB WEB via BOT

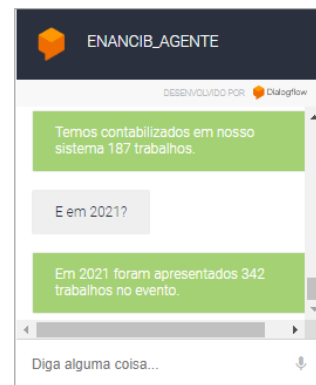


E em 2021?

Em 2021 foram apresentados 342 trabalhos no evento.



ENANCIB WEB via BOT

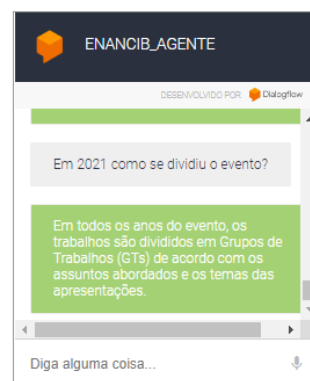


Em 2021 como se dividiu o evento?

Em todos os anos do evento, os trabalhos são divididos em Grupos de Trabalhos (GTs) de acordo com os assuntos abordados e os temas das apresentações.



ENANCIB WEB via BOT



Quais os grupos de trabalhos de 2021?

Em 2021 foram divididos em GTs da seguinte forma:

GT 1 – Estudos Históricos e Epistemológicos da Ciência da Informação

GT 2 – Organização e Representação do Conhecimento

GT 3 – Mediação, Circulação e Apropriação da Informação

GT 4 – Gestão da Informação e do Conhecimento

GT 5 – Política e Economia da Informação

GT 6 – Informação, Educação e Trabalho

GT 7 – Produção e Comunicação da Informação em Ciência, Tecnologia & Inovação

GT 8 – Informação e Tecnologia

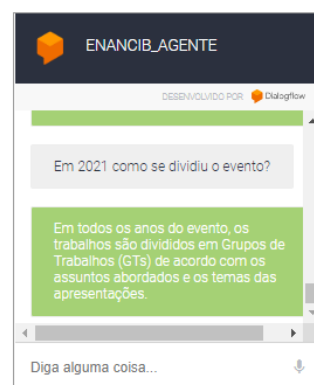
GT 9 – Museu, Patrimônio e Informação

GT 10 – Informação e Memória

GT 11 – Informação & Saúde"



ENANCIB WEB via BOT

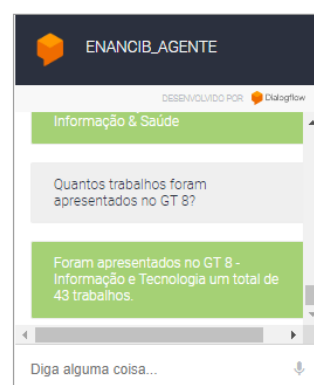


Quantos trabalhos foram apresentados no GT 8?

Foram apresentados no GT 8 - Informação e Tecnologia um total de 43 trabalhos.



ENANCIB WEB via BOT



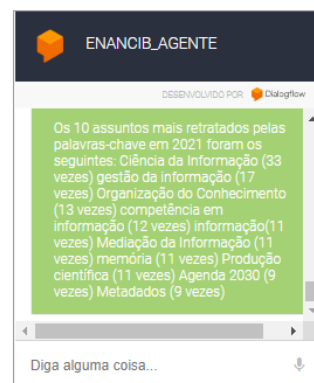
Quais os assuntos das palavras-chave mais apresentados em 2021?

Os 10 assuntos mais retratados pelas palavras-chave em 2021 foram os seguintes:

Ciência da Informação (33 vezes)
gestão da informação (17 vezes)
Organização do Conhecimento (13 vezes)
competência em informação (12 vezes)
informação (11 vezes)
Mediação da Informação (11 vezes)
memória (11 vezes)
Produção científica (11 vezes)
Agenda 2030 (9 vezes)
Metadados (9 vezes)"



ENANCIB WEB via BOT



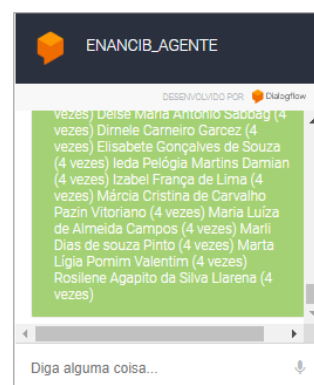
Quais os autores que mais apresentaram trabalhos em 2021?

Os autores que mais apresentaram em 2021 foram os seguintes:

Dalton Lopes Martins (5 vezes)
Gustavo Silva Saldanha (5 vezes)
Silvana Aparecida Borsetti Gregório Vidotti (5 vezes)
Alzira Karla Araújo da Silva (4 vezes)
Carlos Cândido de Almeida (4 vezes)
Cláudio Paixão Anastácio de Paula (4 vezes)
Deise Maria Antonio Sabbag (4 vezes)
Dirnele Carneiro Garcez (4 vezes)
Elisabete Gonçalves de Souza (4 vezes)
Ieda Pelógia Martins Damian (4 vezes)
Izabel França de Lima (4 vezes)
Márcia Cristina de Carvalho Pazin Vitoriano (4 vezes)
Maria Luíza de Almeida Campos (4 vezes)
Marli Dias de souza Pinto (4 vezes)
Marta Lígia Pomim Valentim (4 vezes)
Rosilene Agapito da Silva Llarena (4 vezes)"



ENANCIB WEB via BOT

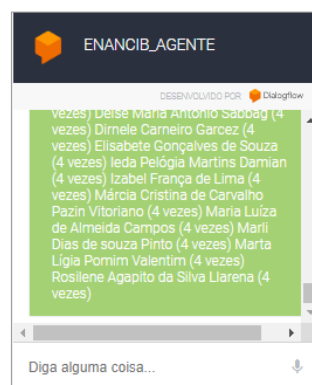


Houve algum trabalho apresentado cuja temática é sobre filmes?

Foram encontrados 15 trabalhos apresentados em diversos anos do evento.



ENANCIB WEB via BOT

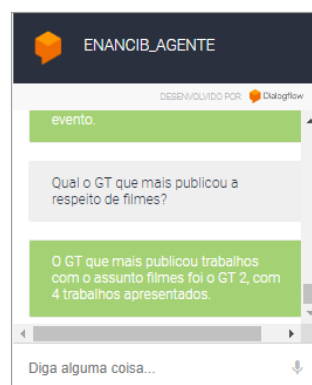


Qual o GT que mais publicou a respeito de filmes?

O GT que mais publicou trabalhos com o assunto filmes foi o GT 2, com 4 trabalhos apresentados.



ENANCIB WEB via BOT

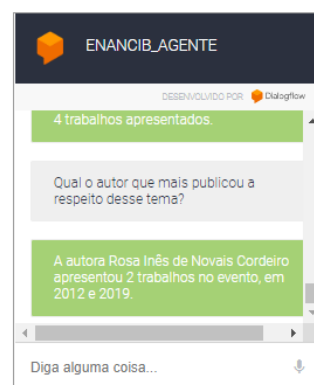


Qual o autor que mais publicou a respeito desse tema?

A autora Rosa Inês de Novais Cordeiro apresentou 2 trabalhos no evento, em 2012 e 2019.



ENANCIB WEB via BOT

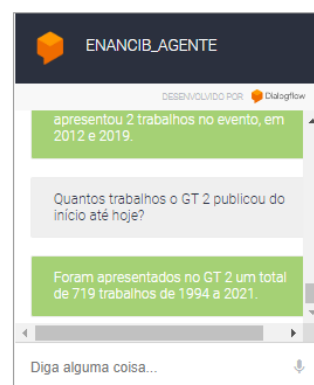


Quantos trabalhos o GT 2 publicou do início até hoje?

Foram apresentados no GT 2 um total de 719 trabalhos de 1994 a 2021.



ENANCIB WEB via BOT



Quantos GTs havia no primeiro ano do evento?

Em 1994, o primeiro ano do evento havia 7 grupos de trabalho:

Informação Tecnológica

Informação e Sociedade/Ação Cultural

Representação do Conhecimento/Indexação/Teoria da Classificação

Administração/Gestão/Avaliação e estudos de usuário

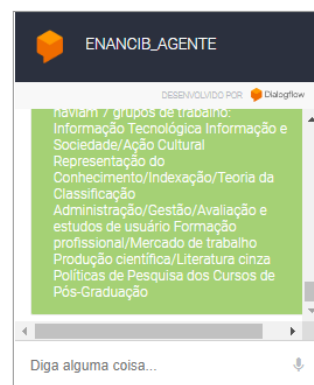
Formação profissional/Mercado de trabalho

Produção científica/Literatura cinza

Políticas de Pesquisa dos Cursos de Pós-Graduação"



ENANCIB WEB via BOT



Quem foi o autor que mais publicou a respeito de tesouros?

Foram encontradas 30 apresentações e os autores que mais apresentaram a respeito do assunto tesouro foram:

Maria Luiza de Almeida Campos (4 vezes)

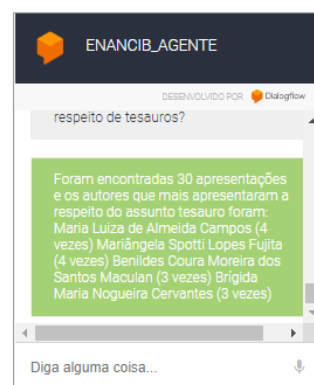
Mariângela Spotti Lopes Fujita (4 vezes)

Benildes Coura Moreira dos Santos Maculan (3 vezes)

Brígida Maria Nogueira Cervantes (3 vezes)"



ENANCIB WEB via BOT



E quem foi o autor que mais publicou a respeito de recuperação de informação?

Foram encontradas 49 apresentações e os autores que mais apresentaram a respeito do assunto recuperação de informação foram:

Renata Maria Abrantes Baracho (5 vezes)

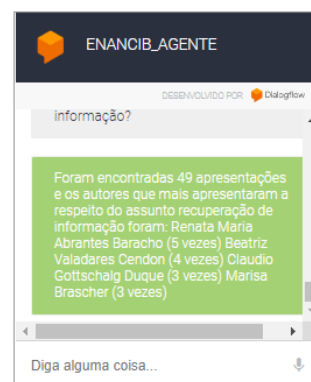
Beatriz Valadares Cendon (4 vezes)

Claudio Gottschalg Duque (3 vezes)

Marisa Brascher (3 vezes)"



ENANCIB WEB via BOT

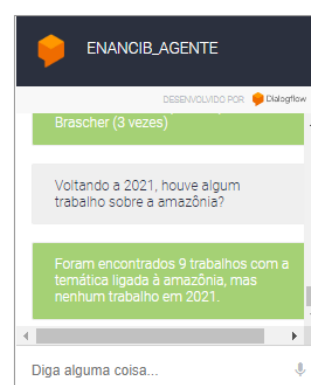


Voltando a 2021, houve algum trabalho sobre a Amazônia?

Foram encontrados 9 trabalhos com a temática ligada à Amazônia, mas nenhum trabalho em 2021.



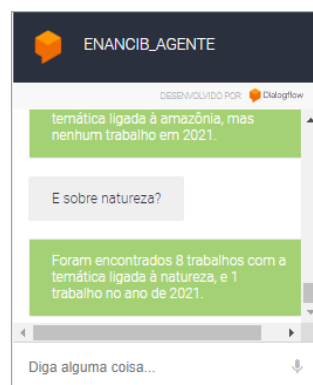
ENANCIB WEB via BOT



E sobre natureza?

Foram encontrados 8 trabalhos com a temática ligada à natureza, e 1 trabalho no ano de 2021.

ENANCIB WEB via BOT



Me mostre este trabalho.

Ano: 2021

GT 1 - Estudos Históricos e Epistemológicos da Ciência da Informação

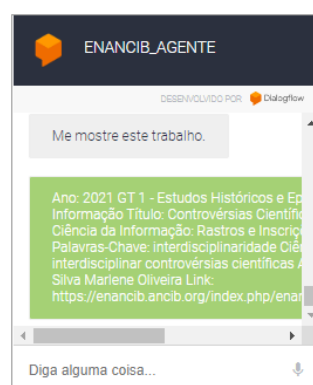
Título:
Controvérsias Científicas da Natureza Interdisciplinar da Ciência da Informação: Rastros e Inscrições de Cientistas da Informação

Palavras-Chave:
interdisciplinaridade
Ciência da Informação
natureza interdisciplinar
controvérsias científicas

Autores:
Zayr Claudio Gomes da Silva
Marlene Oliveira

Link:
<https://enancib.ancib.org/index.php/enancib/xxienancib/paper/view/94/139>

ENANCIB WEB via BOT

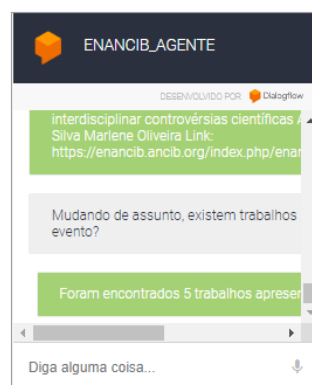


Mudando de assunto, existem trabalhos sobre inteligência artificial nesse evento?

Foram encontrados 5 trabalhos apresentados em diversos anos do evento.



ENANCIB WEB via BOT



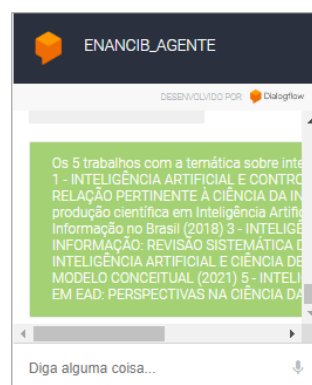
Me mostre os títulos.

"Os 5 trabalhos com a temática sobre inteligência artificial encontrados são:

- 1 - INTELIGÊNCIA ARTIFICIAL E CONTROLE SOCIAL DA CT&I: UMA RELAÇÃO PERTINENTE À CIÊNCIA DA INFORMAÇÃO (2014)
- 2 - Análise da produção científica em Inteligência Artificial na área da Ciência da Informação no Brasil (2018)
- 3 - INTELIGÊNCIA ARTIFICIAL EM CIÊNCIA DA INFORMAÇÃO: REVISÃO SISTEMÁTICA DA LITERATURA (2021)
- 4 - INTELIGÊNCIA ARTIFICIAL E CIÊNCIA DE DADOS EM CRIS INSTITUCIONAL: MODELO CONCEITUAL (2021)
- 5 - INTELIGÊNCIA ARTIFICIAL NA TUTORIA EM EAD: PERSPECTIVAS NA CIÊNCIA DA INFORMAÇÃO (2021)"



ENANCIB WEB via BOT



Me mostre o trabalho 3Ano: 2021

GT 8 - Informação e Tecnologia

Título:
INTELIGÊNCIA ARTIFICIAL EM CIÊNCIA DA
INFORMAÇÃO: REVISÃO SISTEMÁTICA DA
LITERATURA

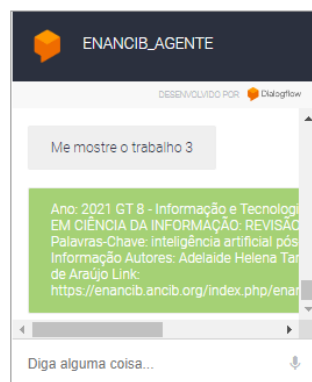
Palavras-Chave:
inteligência artificial
pós-humanismo
Ciência da Informação

Autores:
Adelaide Helena Targino Casimiro
Wagner Junqueira de Araújo

Link:
<https://enancib.ancib.org/index.php/enancib/xxienancib/paper/view/8/442>



ENANCIB WEB via BOT



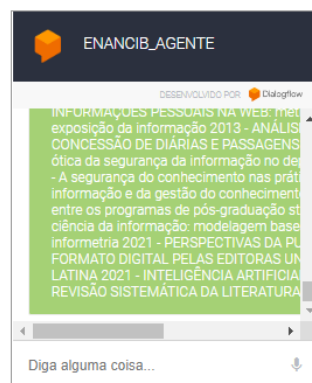
Estes autores já publicaram em outros anos do evento?

Outras publicações destes autores são:
Adelaide Helena Targino Casimiro
2018 - GESTÃO DO CONHECIMENTO E
PRÁTICAS ARQUIVÍSTICAS: CONSTRUTOS
TEÓRICOS PARA FORMAÇÃO E
PRESERVAÇÃO DA MEMÓRIA
ORGANIZACIONAL
2021 - MEDIAÇÃO DA INFORMAÇÃO E
COMPETÊNCIA EM INFORMAÇÃO: UMA
RELAÇÃO POSSÍVEL EM CENÁRIO
PANDÊMICO
2021 - INTELIGÊNCIA ARTIFICIAL EM
CIÊNCIA DA INFORMAÇÃO: REVISÃO
SISTEMÁTICA DA LITERATURA

Wagner Junqueira de Araújo
2019 - PROSPECÇÃO DE CENÁRIOS: ESTUDO
SOBRE A IMPLANTAÇÃO DO PROCESSO
ELETRÔNICO NACIONAL NA UNIVERSIDADE
FEDERAL DA PARAÍBA
2018 - VALIDAÇÃO DO MAPEAMENTO DE
FLUXOS DE INFORMAÇÃO EM PROCESSOS
ORGANIZACIONAIS: UMA ABORDAGEM COM
FOCO ARQUIVÍSTICO
2018 - CIBER TERRORISMO NA PARAÍBA
2017 - ESTUDO DE CENÁRIOS PROSPECTIVOS
PARA O MERCADO EDITORIAL DO LIVRO
DIGITAL NO BRASIL
2017 - SITES DOS MUNICÍPIOS DA PARAÍBA:
ANÁLISE DE VULNERABILIDADES
COMPUTACIONAIS



ENANCIB WEB via BOT



2016 - ANÁLISE DA DIMENSÃO HUMANA NO PROCESSO DE GESTÃO DE SEGURANÇA DA INFORMAÇÃO

2016 - APLICAÇÃO DO OPEN MONOGRAPH PRESS POR EDITORAS BRASILEIRAS

2015 - ASPECTOS HUMANOS NA GESTÃO DA SEGURANÇA DA INFORMAÇÃO

2015 - Aspectos de Vulnerabilidades em Bibliotecas Digitais Acessíveis

2014 - ANÁLISE DO SISTEMA DE INFORMAÇÃO DA BIBLIOTECA CENTRAL DA UFPB NOS PROCESSOS DE GESTÃO DA INFORMAÇÃO PARA O SETOR DE REFERÊNCIA

2013 - ANÁLISE DE INFORMAÇÕES PESSOAIS NA WEB: métrica para identificar o grau de exposição da informação

2013 - ANÁLISE DE RISCO NO SISTEMA DE CONCESSÃO DE DIÁRIAS E PASSAGENS (SCDP): estudo de caso sob a ótica da segurança da informação no departamento contábil da UFPB

2010 - A segurança do conhecimento nas práticas da gestão da segurança da informação e da gestão do conhecimento

2010 - Relações de colaboração entre os programas de pós-graduação stricto sensu brasileiros na área da ciência da informação: modelagem baseada em grafos e programa de informetria

2021 - PERSPECTIVAS DA PUBLICAÇÃO DE LIVROS EM FORMATO DIGITAL PELAS EDITORAS UNIVERSITÁRIAS NA ÁMÉRICA LATINA

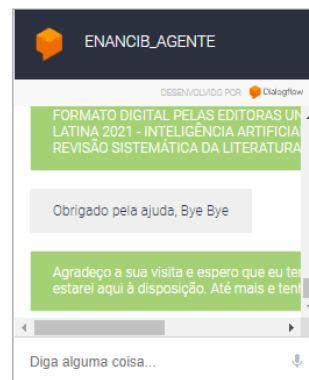
2021 - INTELIGÊNCIA ARTIFICIAL EM CIÊNCIA DA INFORMAÇÃO: REVISÃO SISTEMÁTICA DA LITERATURA"

Obrigado pela ajuda, Bye Bye

Agradeço a sua visita e espero que eu tenha ajudado. Sempre que precisar estarei aqui à disposição. Até mais e tenha uma Boa Semana.

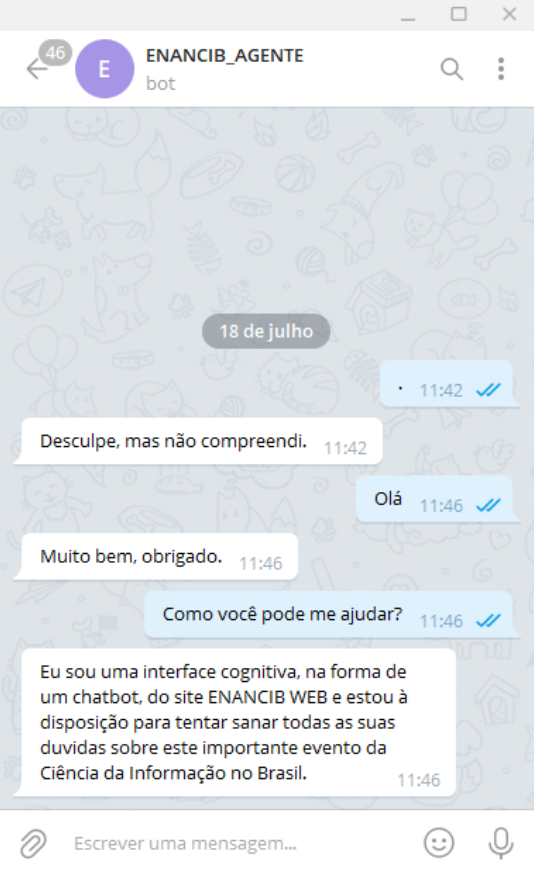


ENANCIB WEB via BOT



ANEXO II

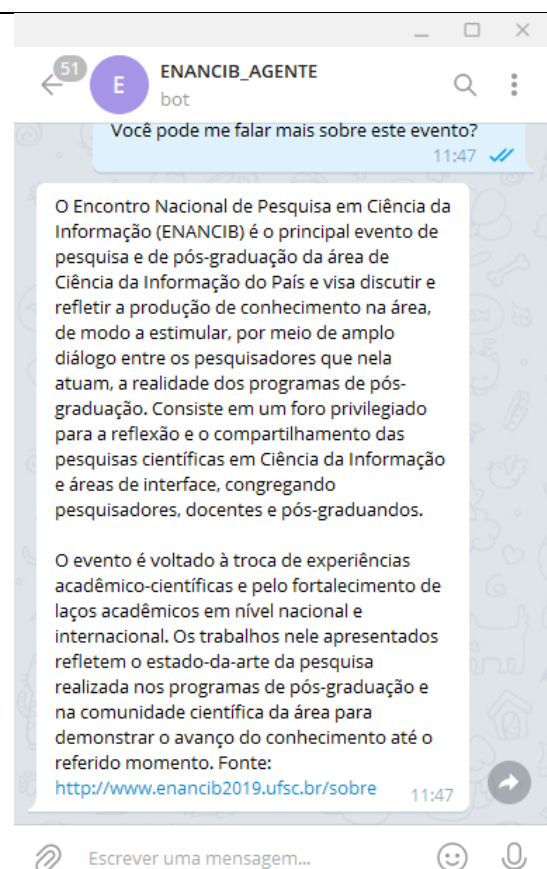
Neste Anexo são mostradas as interações efetuadas por meio do *software* de mensagens Telegram instalado em um computador de mesa com Windows 11 e respondendo às perguntas de um usuário, neste caso por texto.

TEXTOS DOS DIÁLOGOS	IMAGENS DOS DIÁLOGOS VIA TELEGRAM DESKTOP
Olá Muito bem, obrigado.	
Como você pode me ajudar? Eu sou uma interface cognitiva, na forma de um chatbot, do site ENANCIB WEB e estou à disposição para tentar sanar todas as suas dúvidas sobre este importante evento da Ciência da Informação no Brasil.	

Você pode me falar mais sobre este evento?

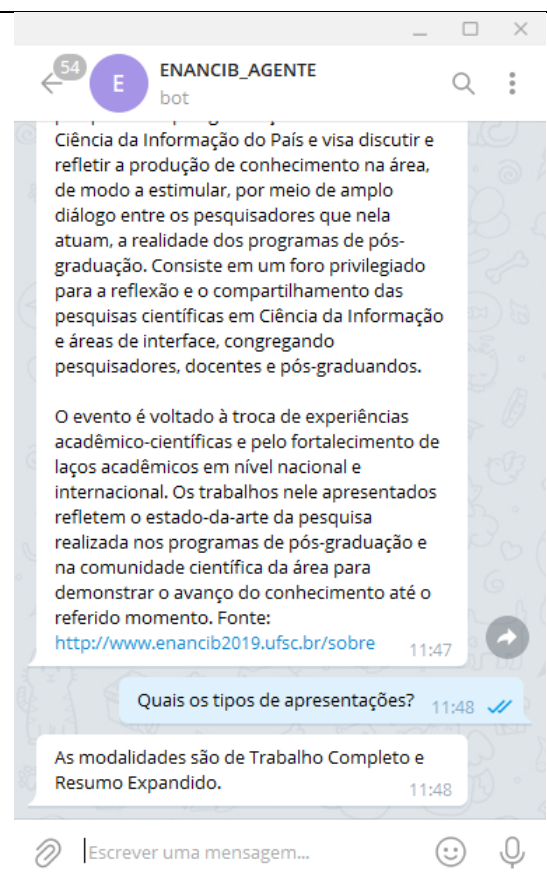
O Encontro Nacional de Pesquisa em Ciência da Informação (ENANCIB) é o principal evento de pesquisa e de pós-graduação da área de Ciência da Informação do País e visa discutir e refletir a produção de conhecimento na área, de modo a estimular, por meio de amplo diálogo entre os pesquisadores que nela atuam, a realidade dos programas de pós-graduação. Consiste em um foro privilegiado para a reflexão e o compartilhamento das pesquisas científicas em Ciência da Informação e áreas de interface, congregando pesquisadores, docentes e pós-graduandos.

O evento é voltado à troca de experiências acadêmico-científicas e pelo fortalecimento de laços acadêmicos em nível nacional e internacional. Os trabalhos nele apresentados refletem o estado-da-arte da pesquisa realizada nos programas de pós-graduação e na comunidade científica da área para demonstrar o avanço do conhecimento até o referido momento. Fonte: <http://www.enancib2019.ufsc.br/sobre>



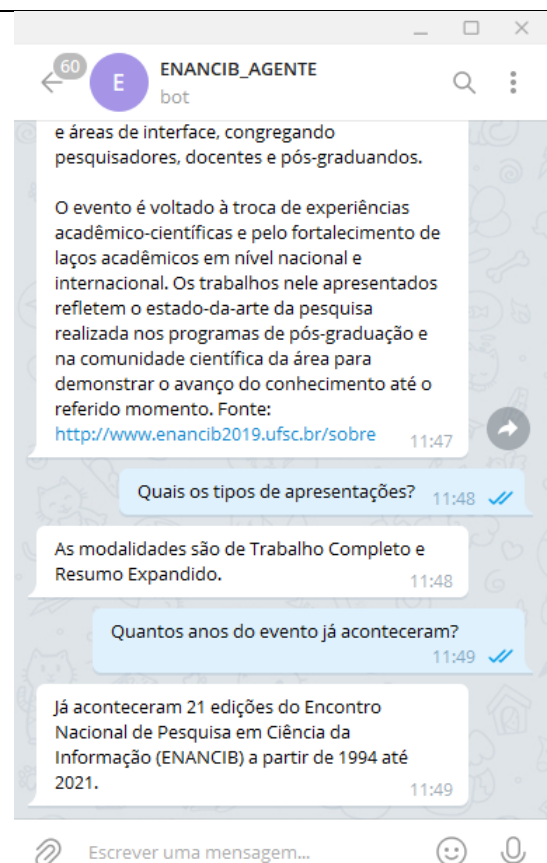
Quais os tipos de apresentações?

As modalidades são de Trabalho Completo e Resumo Expandido.



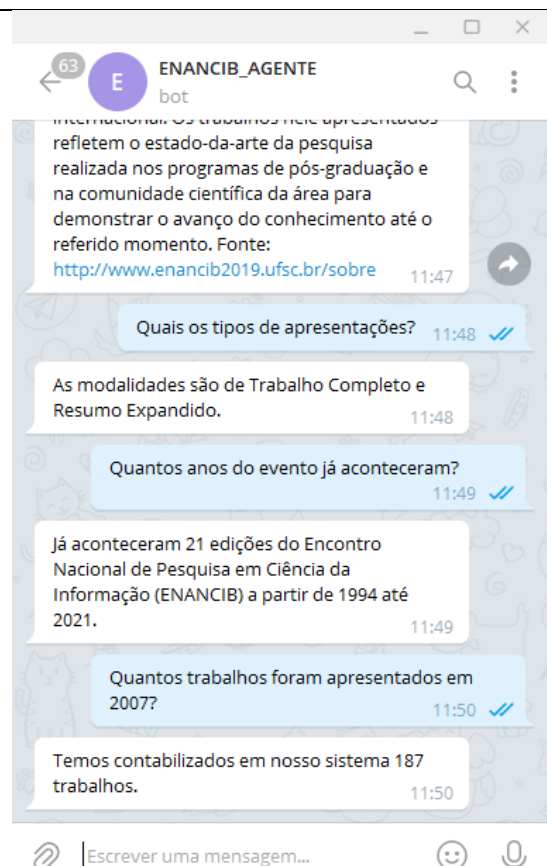
Quanto anos do evento já aconteceram?

Já aconteceram 21 edições do Encontro Nacional de Pesquisa em Ciência da Informação (ENANCIB) a partir de 1994 até 2021.



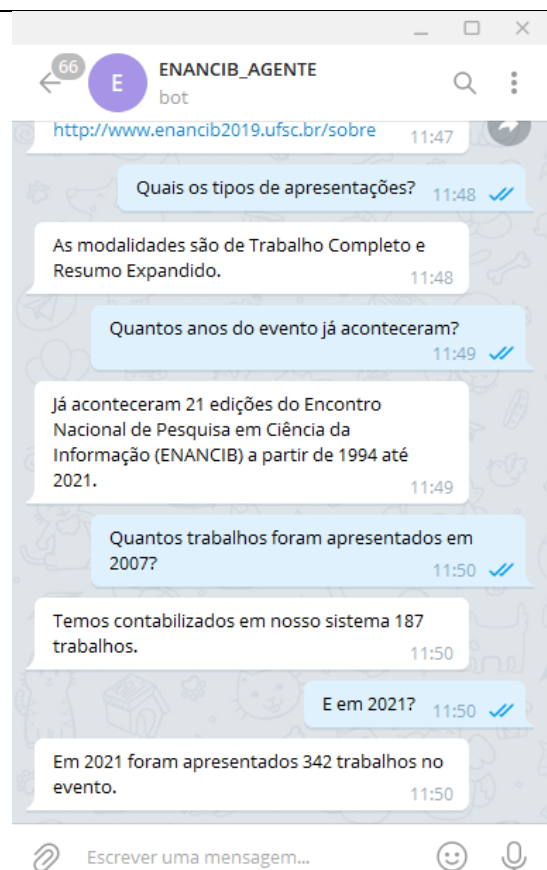
Quanto trabalhos foram apresentados em 2007?

Temos contabilizados em nosso sistema 187 trabalhos.



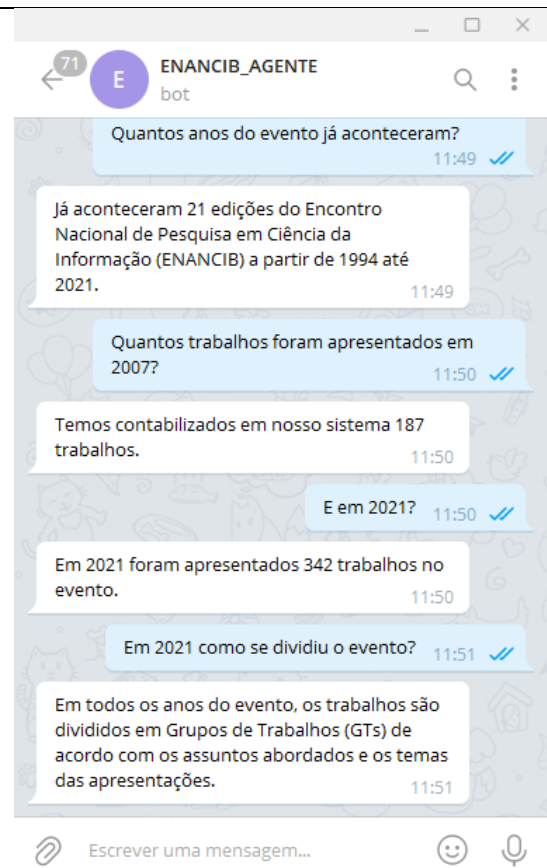
E em 2021?

Em 2021 foram apresentados 342 trabalhos no evento.



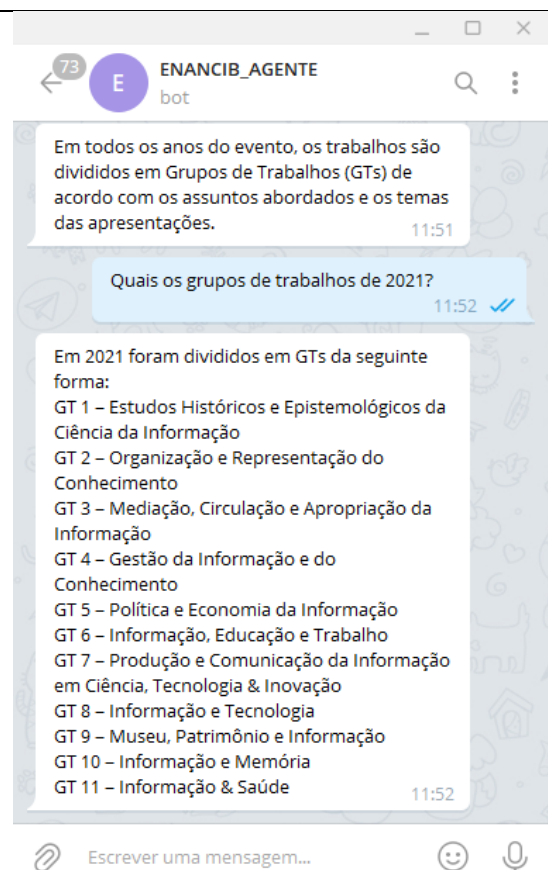
Em 2021 como se dividiu o evento?

Em todos os anos do evento, os trabalhos são divididos em Grupos de Trabalhos (GTs) de acordo com os assuntos abordados e os temas das apresentações.



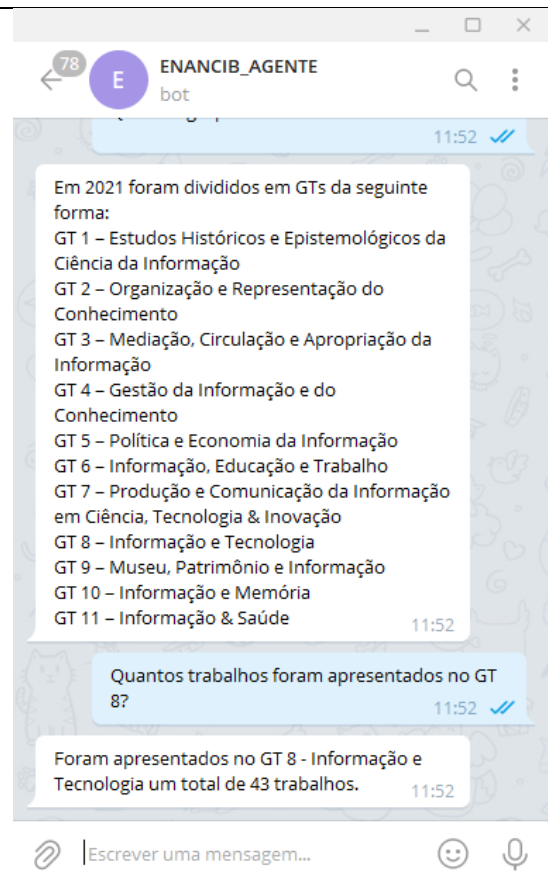
Quais os grupos de trabalhos de 2021?

Em 2021 foram divididos em GTs da seguinte forma:
GT 1 – Estudos Históricos e Epistemológicos da Ciência da Informação
GT 2 – Organização e Representação do Conhecimento
GT 3 – Mediação, Circulação e Apropriação da Informação
GT 4 – Gestão da Informação e do Conhecimento
GT 5 – Política e Economia da Informação
GT 6 – Informação, Educação e Trabalho
GT 7 – Produção e Comunicação da Informação em Ciência, Tecnologia & Inovação
GT 8 – Informação e Tecnologia
GT 9 – Museu, Patrimônio e Informação
GT 10 – Informação e Memória
GT 11 – Informação & Saúde"



Quantos trabalhos foram apresentados no GT 8?

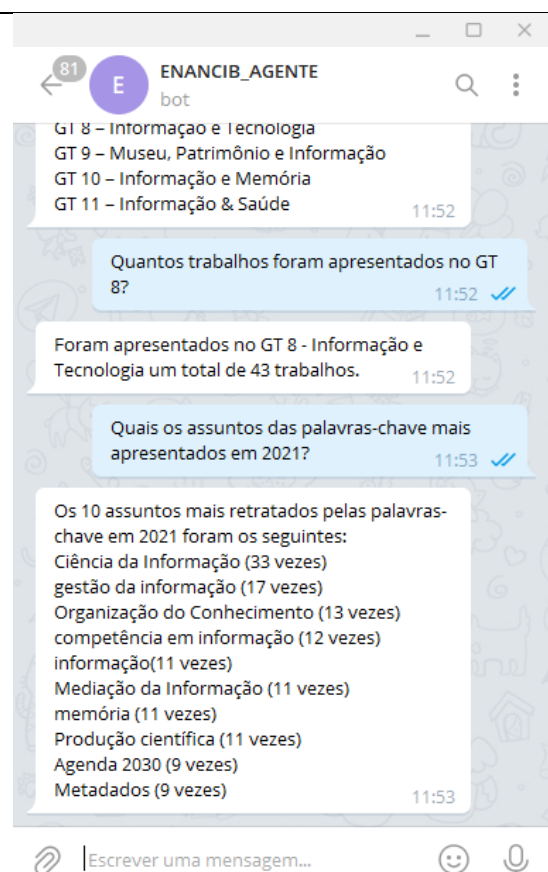
Foram apresentados no GT 8 - Informação e Tecnologia um total de 43 trabalhos.



Quais os assuntos das palavras-chave mais apresentados em 2021?

Os 10 assuntos mais retratados pelas palavras-chave em 2021 foram os seguintes:

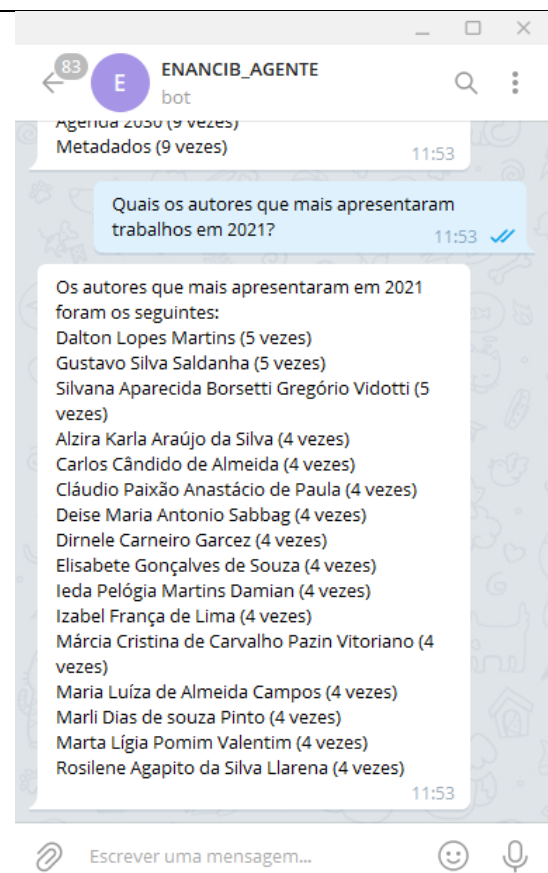
Ciência da Informação (33 vezes)
gestão da informação (17 vezes)
Organização do Conhecimento (13 vezes)
competência em informação (12 vezes)
informação (11 vezes)
Mediação da Informação (11 vezes)
memória (11 vezes)
Produção científica (11 vezes)
Agenda 2030 (9 vezes)
Metadados (9 vezes)"



Quais os autores que mais apresentaram trabalhos em 2021?

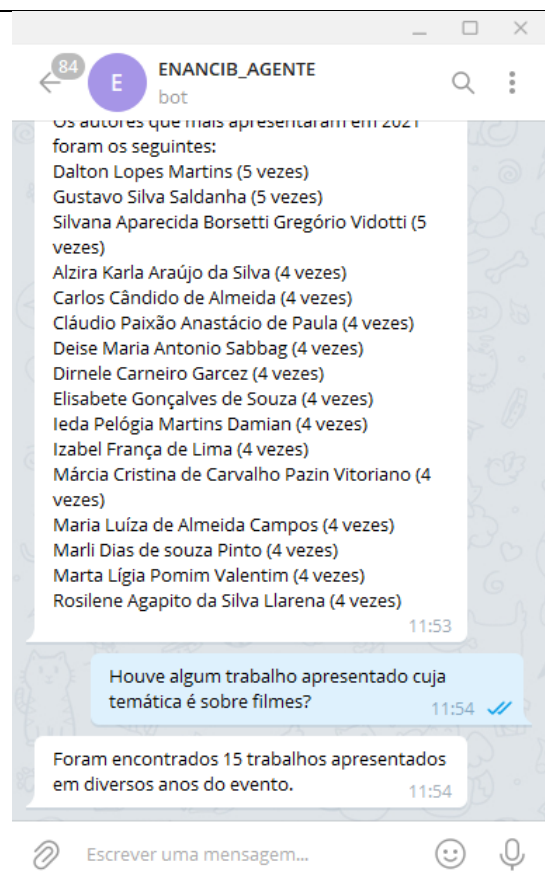
Os autores que mais apresentaram em 2021 foram os seguintes:

Dalton Lopes Martins (5 vezes)
Gustavo Silva Saldanha (5 vezes)
Silvana Aparecida Borsetti Gregório Vidotti (5 vezes)
Alzira Karla Araújo da Silva (4 vezes)
Carlos Cândido de Almeida (4 vezes)
Cláudio Paixão Anastácio de Paula (4 vezes)
Deise Maria Antonio Sabbag (4 vezes)
Dirnele Carneiro Garcez (4 vezes)
Elisabete Gonçalves de Souza (4 vezes)
Ieda Pelógia Martins Damian (4 vezes)
Izabel França de Lima (4 vezes)
Márcia Cristina de Carvalho Pazin Vitoriano (4 vezes)
Maria Luíza de Almeida Campos (4 vezes)
Marli Dias de souza Pinto (4 vezes)
Marta Lúgia Pomim Valentim (4 vezes)
Rosilene Agapito da Silva Llarena (4 vezes)"



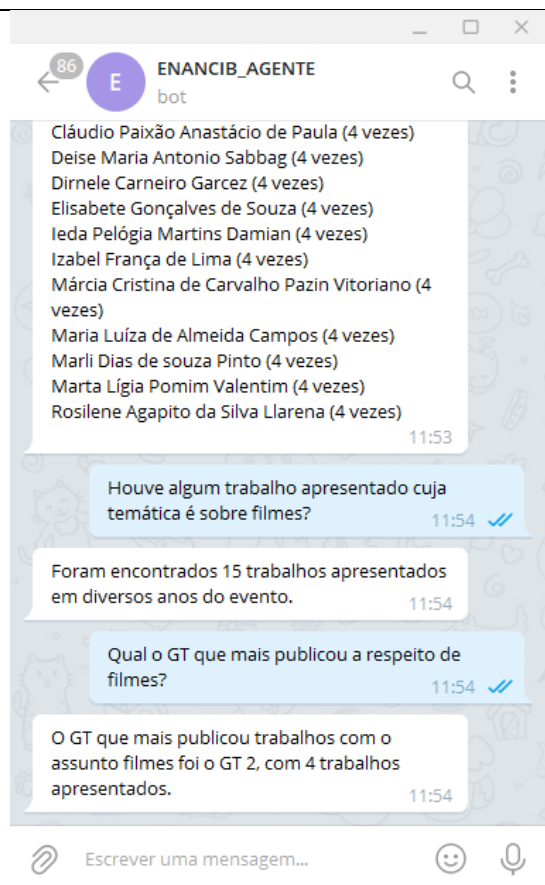
Houve algum trabalho apresentado cuja temática é sobre filmes?

Foram encontrados 15 trabalhos apresentados em diversos anos do evento.



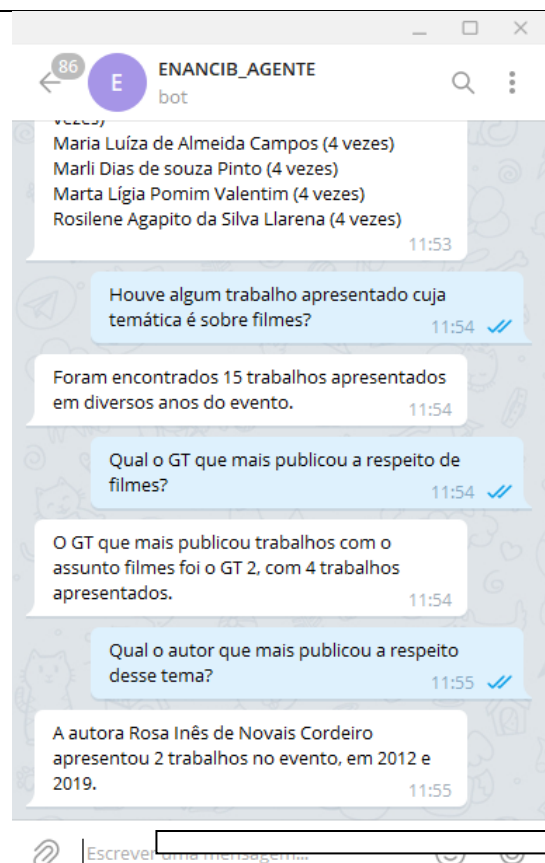
Qual o GT que mais publicou a respeito de filmes?

O GT que mais publicou trabalhos com o assunto filmes foi o GT 2, com 4 trabalhos apresentados.



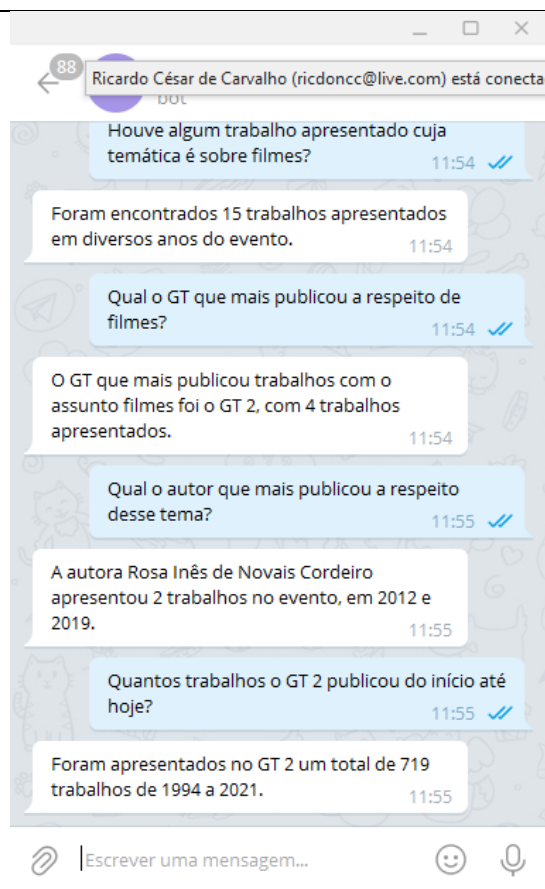
Qual o autor que mais publicou a respeito desse tema?

A autora Rosa Inês de Novais Cordeiro apresentou 2 trabalhos no evento, em 2012 e 2019.



Quantos trabalhos o GT 2 publicou do início até hoje?

Foram apresentados no GT 2 um total de 719 trabalhos de 1994 a 2021.



Quantos GTs havia no primeiro ano do evento?

Em 1994, o primeiro ano do evento havia 7 grupos de trabalho:

Informação Tecnológica

Informação e Sociedade/Ação Cultural

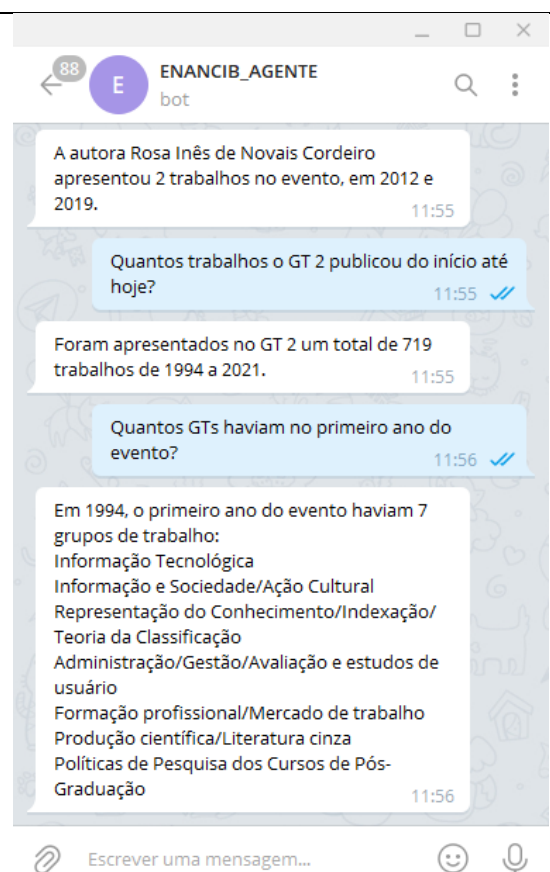
Representação do Conhecimento/Indexação/Teoria da Classificação

Administração/Gestão/Avaliação e estudos de usuário

Formação profissional/Mercado de trabalho

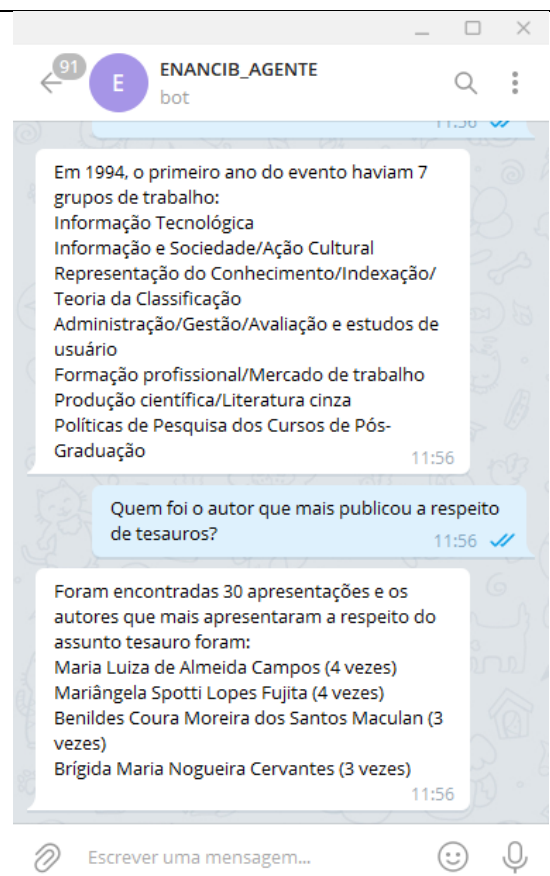
Produção científica/Literatura cinza

Políticas de Pesquisa dos Cursos de Pós-Graduação



Quem foi o autor que mais publicou a respeito de tesouros?

Foram encontradas 30 apresentações e os autores que mais apresentaram a respeito do assunto tesouro foram: Maria Luiza de Almeida Campos (4 vezes)
Mariângela Spotti Lopes Fujita (4 vezes)
Benildes Coura Moreira dos Santos Maculan (3 vezes)
Brígida Maria Nogueira Cervantes (3 vezes)"



E quem foi o autor que mais publicou a respeito de recuperação de informação?

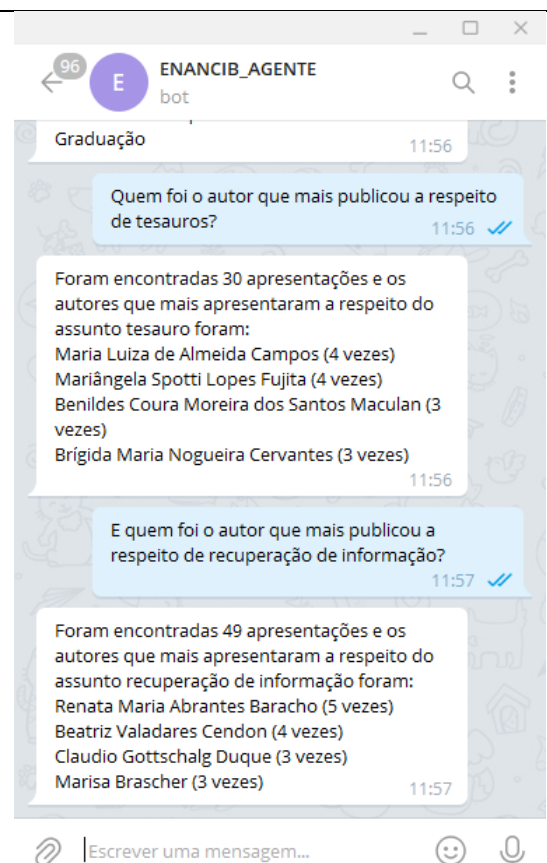
Foram encontradas 49 apresentações e os autores que mais apresentaram a respeito do assunto recuperação de informação foram:

Renata Maria Abrantes Baracho (5 vezes)

Beatriz Valadares Cendon (4 vezes)

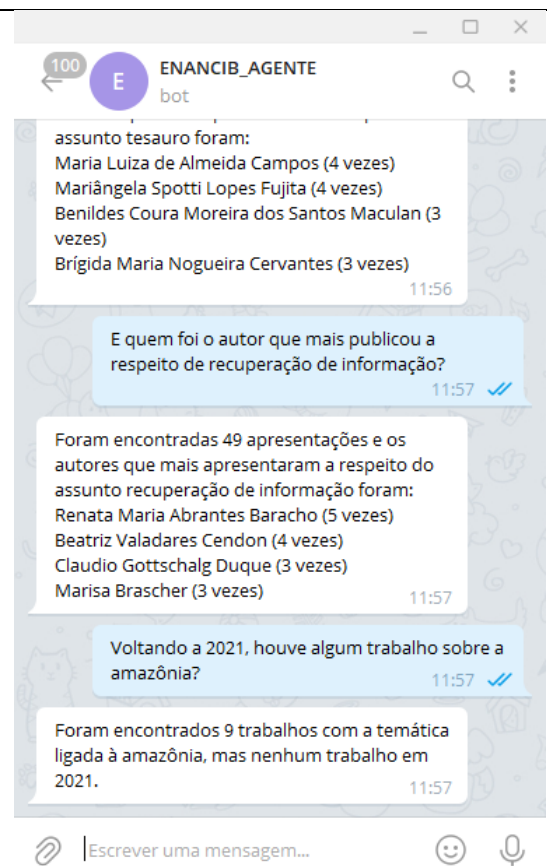
Claudio Gottschalg Duque (3 vezes)

Marisa Brascher (3 vezes)"



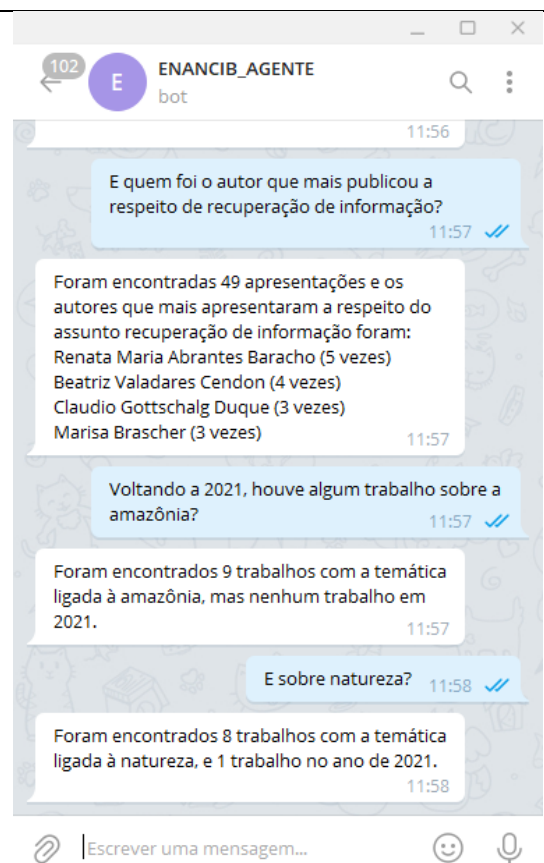
Voltando a 2021, houve algum trabalho sobre a Amazônia?

Foram encontrados 9 trabalhos com a temática ligada à Amazônia, mas nenhum trabalho em 2021.



E sobre natureza?

Foram encontrados 8 trabalhos com a temática ligada à natureza, e 1 trabalho no ano de 2021.



Me mostre este trabalho.

Ano: 2021

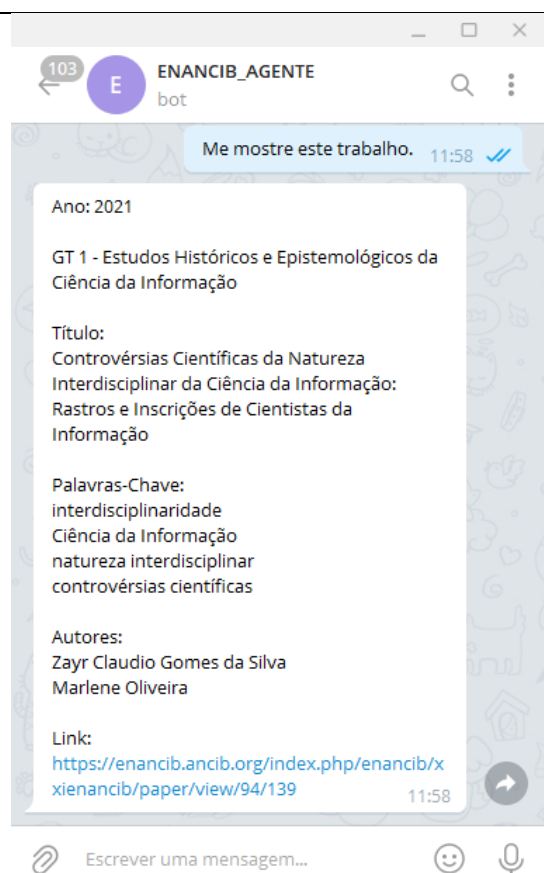
GT 1 - Estudos Históricos e Epistemológicos da Ciência da Informação

Título:
Controvérsias Científicas da Natureza Interdisciplinar da Ciência da Informação: Rastros e Inscrições de Cientistas da Informação

Palavras-Chave:
interdisciplinaridade
Ciência da Informação
natureza interdisciplinar
controvérsias científicas

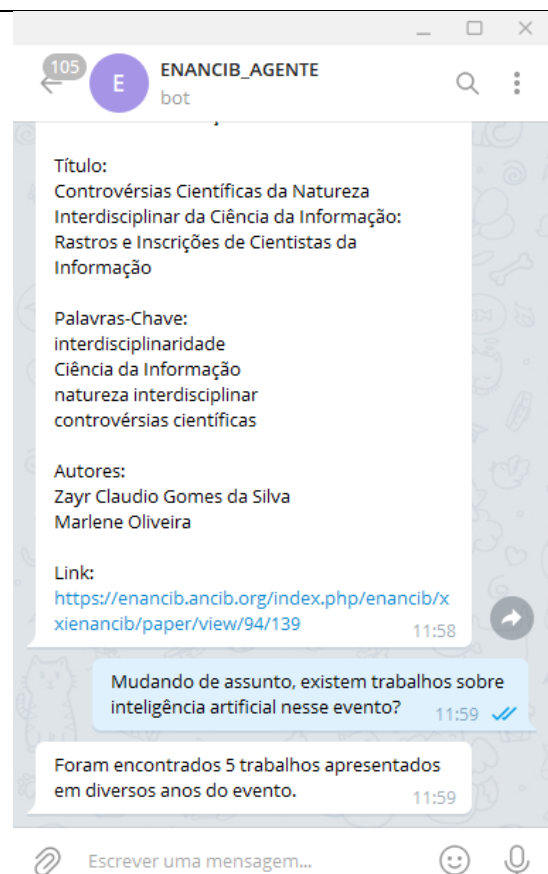
Autores:
Zayr Claudio Gomes da Silva
Marlene Oliveira

Link:
<https://enancib.ancib.org/index.php/enancib/xxienancib/paper/view/94/139>



Mudando de assunto, existem trabalhos sobre inteligência artificial nesse evento?

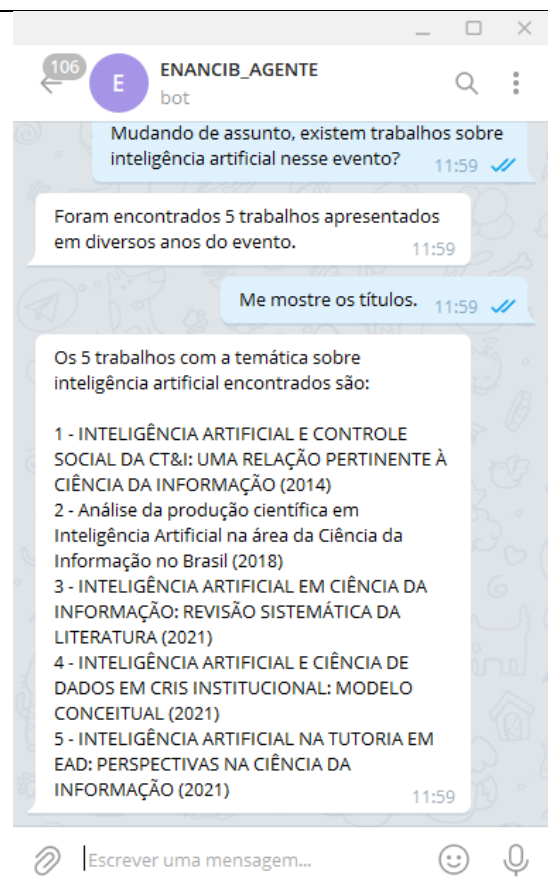
Foram encontrados 5 trabalhos apresentados em diversos anos do evento.



Me mostre os títulos.

Os 5 trabalhos com a temática sobre inteligência artificial encontrados são:

- 1 - INTELIGÊNCIA ARTIFICIAL E CONTROLE SOCIAL DA CT&I: UMA RELAÇÃO PERTINENTE À CIÊNCIA DA INFORMAÇÃO (2014)
- 2 - Análise da produção científica em Inteligência Artificial na área da Ciência da Informação no Brasil (2018)
- 3 - INTELIGÊNCIA ARTIFICIAL EM CIÊNCIA DA INFORMAÇÃO: REVISÃO SISTEMÁTICA DA LITERATURA (2021)
- 4 - INTELIGÊNCIA ARTIFICIAL E CIÊNCIA DE DADOS EM CRIS INSTITUCIONAL: MODELO CONCEITUAL (2021)
- 5 - INTELIGÊNCIA ARTIFICIAL NA TUTORIA EM EAD: PERSPECTIVAS NA CIÊNCIA DA INFORMAÇÃO (2021)"



Me mostre o trabalho 3

Ano: 2021

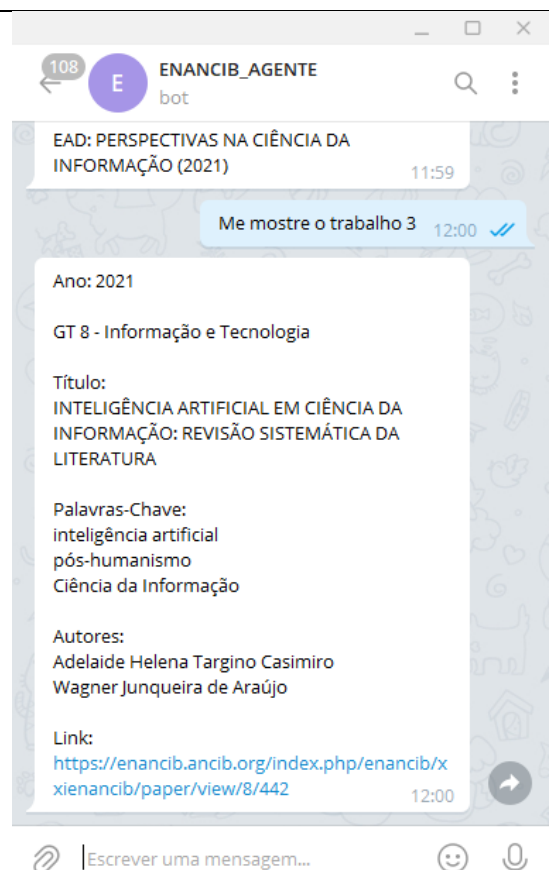
GT 8 - Informação e Tecnologia

Título:
INTELIGÊNCIA ARTIFICIAL EM CIÊNCIA DA
INFORMAÇÃO: REVISÃO SISTEMÁTICA DA
LITERATURA

Palavras-Chave:
inteligência artificial
pós-humanismo
Ciência da Informação

Autores:
Adelaide Helena Targino Casimiro
Wagner Junqueira de Araújo

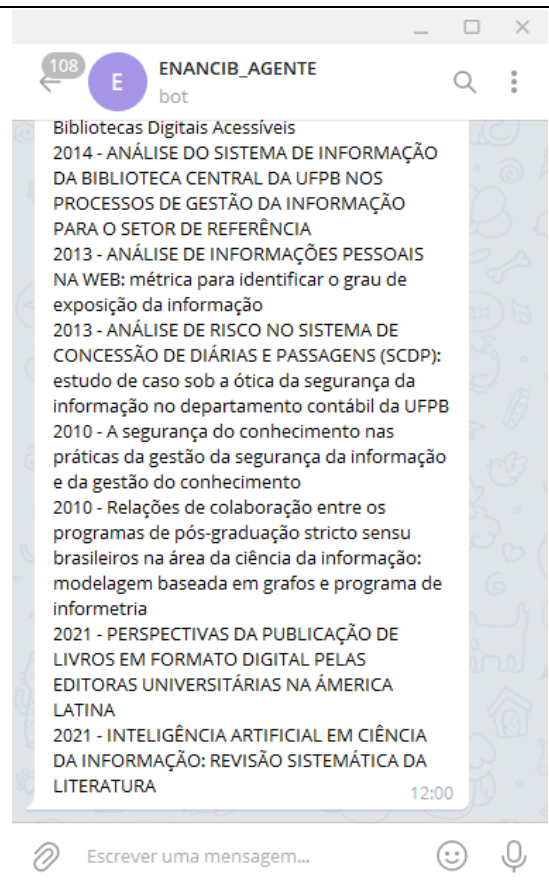
Link:
<https://enancib.ancib.org/index.php/enancib/xxienancib/paper/view/8/442>



Estes autores já publicaram em outros anos do evento?

Outras publicações destes autores são:
Adelaide Helena Targino Casimiro
2018 - GESTÃO DO CONHECIMENTO E PRÁTICAS ARQUIVÍSTICAS: CONSTRUTOS TEÓRICOS PARA FORMAÇÃO E PRESERVAÇÃO DA MEMÓRIA ORGANIZACIONAL
2021 - MEDIAÇÃO DA INFORMAÇÃO E COMPETÊNCIA EM INFORMAÇÃO: UMA RELAÇÃO POSSÍVEL EM CENÁRIO PANDÊMICO
2021 - INTELIGÊNCIA ARTIFICIAL EM CIÊNCIA DA INFORMAÇÃO: REVISÃO SISTEMÁTICA DA LITERATURA

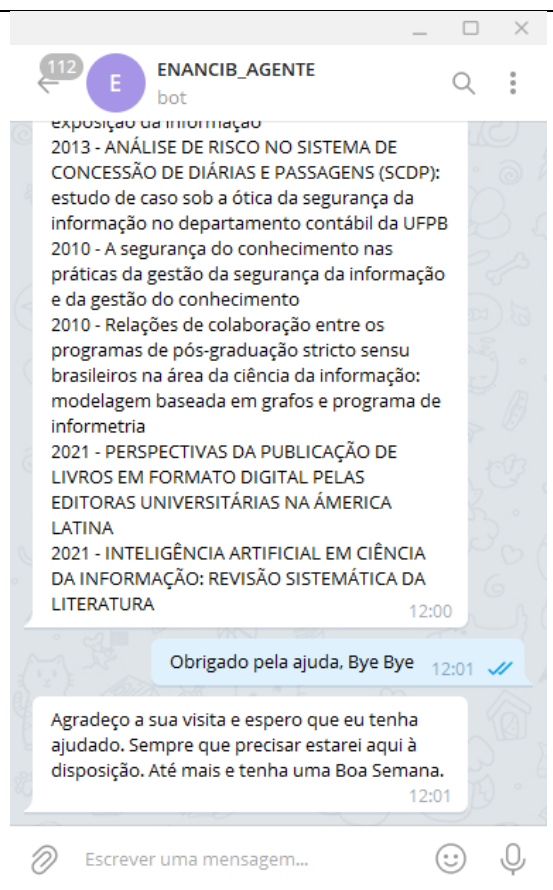
Wagner Junqueira de Araújo
2019 - PROSPECÇÃO DE CENÁRIOS: ESTUDO SOBRE A IMPLANTAÇÃO DO PROCESSO ELETRÔNICO NACIONAL NA UNIVERSIDADE FEDERAL DA PARAÍBA
2018 - VALIDAÇÃO DO MAPEAMENTO DE FLUXOS DE INFORMAÇÃO EM PROCESSOS ORGANIZACIONAIS: UMA ABORDAGEM COM FOCO ARQUIVÍSTICO
2018 - CIBER TERRORISMO NA PARAÍBA
2017 - ESTUDO DE CENÁRIOS PROSPECTIVOS PARA O MERCADO EDITORIAL DO LIVRO DIGITAL NO BRASIL



2017 - SITES DOS MUNICÍPIOS DA PARAÍBA: ANÁLISE DE VULNERABILIDADES COMPUTACIONAIS 2016 - ANÁLISE DA DIMENSÃO HUMANA NO PROCESSO DE GESTÃO DE SEGURANÇA DA INFORMAÇÃO 2016 - APLICAÇÃO DO OPEN MONOGRAPH PRESS POR EDITORAS BRASILEIRAS 2015 - ASPECTOS HUMANOS NA GESTÃO DA SEGURANÇA DA INFORMAÇÃO 2015 - Aspectos de Vulnerabilidades em Bibliotecas Digitais Acessíveis 2014 - ANÁLISE DO SISTEMA DE INFORMAÇÃO DA BIBLIOTECA CENTRAL DA UFPB NOS PROCESSOS DE GESTÃO DA INFORMAÇÃO PARA O SETOR DE REFERÊNCIA 2013 - ANÁLISE DE INFORMAÇÕES PESSOAIS NA WEB: métrica para identificar o grau de exposição da informação 2013 - ANÁLISE DE RISCO NO SISTEMA DE CONCESSÃO DE DIÁRIAS E PASSAGENS (SCDP): estudo de caso sob a ótica da segurança da informação no departamento contábil da UFPB 2010 - A segurança do conhecimento nas práticas da gestão da segurança da informação e da gestão do conhecimento 2010 - Relações de colaboração entre os programas de pós-graduação stricto sensu brasileiros na área da ciência da informação: modelagem baseada em grafos e programa de informetria 2021 - PERSPECTIVAS DA PUBLICAÇÃO DE LIVROS EM FORMATO DIGITAL PELAS EDITORAS UNIVERSITÁRIAS NA AMÉRICA LATINA 2021 - INTELIGÊNCIA ARTIFICIAL EM CIÊNCIA DA INFORMAÇÃO: REVISÃO SISTEMÁTICA DA LITERATURA"	
---	--

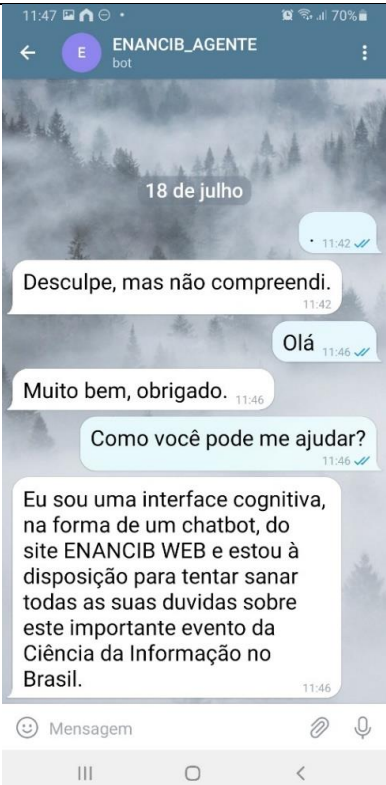
Obrigado pela ajuda, Bye Bye

Agradeço a sua visita e espero que eu tenha ajudado. Sempre que precisar estarei aqui à disposição. Até mais e tenha uma Boa Semana.



ANEXO III

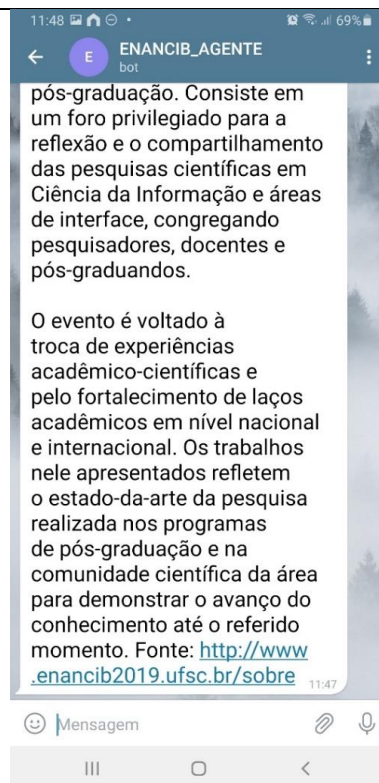
Neste Anexo são mostradas as interações efetuadas por meio do *software* de mensagens Telegram instalado em um celular Android respondendo às perguntas de um usuário, neste caso por texto.

TEXTOS DOS DIÁLOGOS	IMAGENS DOS DIÁLOGOS VIA TELEGRAM NO CELULAR
Olá Muito bem, obrigado.	
Como você pode me ajudar? Eu sou uma interface cognitiva, na forma de um chatbot, do site ENANCIB WEB e estou à disposição para tentar sanar todas as suas dúvidas sobre este importante evento da Ciência da Informação no Brasil.	

Você pode me falar mais sobre este evento?

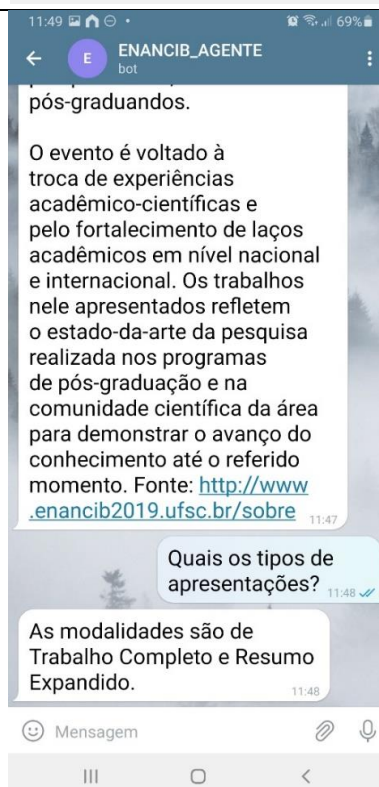
O Encontro Nacional de Pesquisa em Ciência da Informação (ENANCIB) é o principal evento de pesquisa e de pós-graduação da área de Ciência da Informação do País e visa discutir e refletir a produção de conhecimento na área, de modo a estimular, por meio de amplo diálogo entre os pesquisadores que nela atuam, a realidade dos programas de pós-graduação. Consiste em um foro privilegiado para a reflexão e o compartilhamento das pesquisas científicas em Ciência da Informação e áreas de interface, congregando pesquisadores, docentes e pós-graduandos.

O evento é voltado à troca de experiências acadêmico-científicas e pelo fortalecimento de laços acadêmicos em nível nacional e internacional. Os trabalhos nele apresentados refletem o estado-da-arte da pesquisa realizada nos programas de pós-graduação e na comunidade científica da área para demonstrar o avanço do conhecimento até o referido momento. Fonte: <http://www.enancib2019.ufsc.br/sobre>



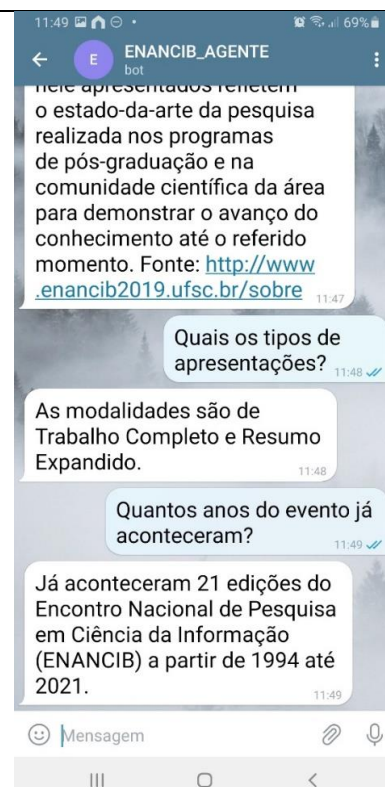
Quais os tipos de apresentações?

As modalidades são de Trabalho Completo e Resumo Expandido.



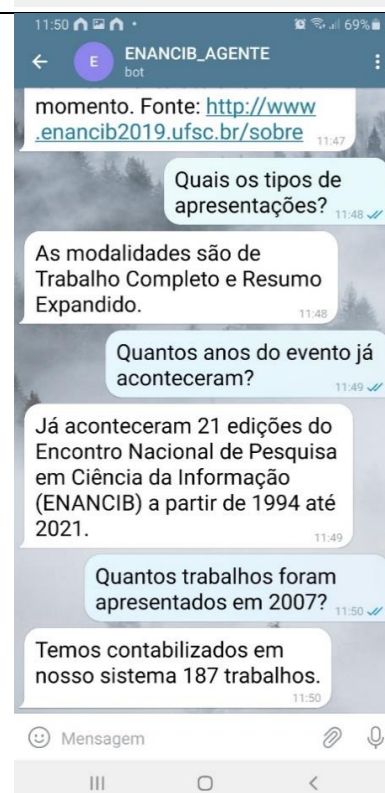
Quantos anos do evento já aconteceram?

Já aconteceram 21 edições do Encontro Nacional de Pesquisa em Ciência da Informação (ENANCIB) a partir de 1994 até 2021.



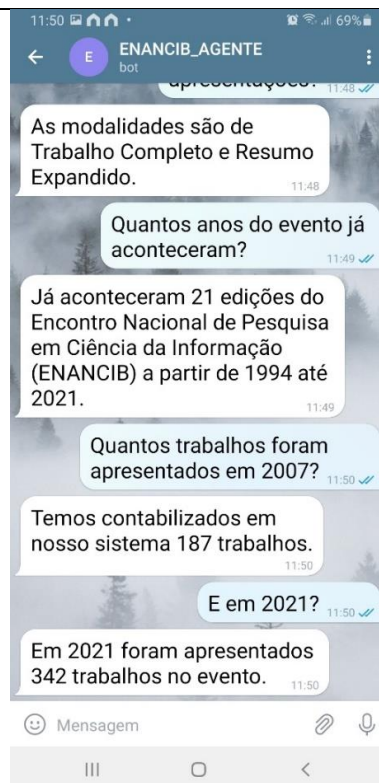
Quantos trabalhos foram apresentados em 2007?

Temos contabilizados em nosso sistema 187 trabalhos.



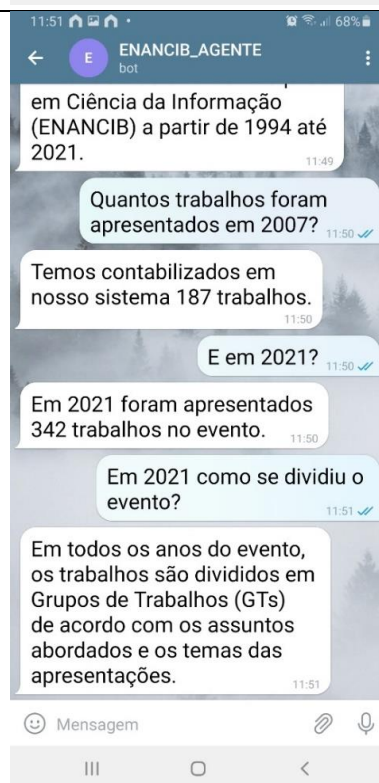
E em 2021?

Em 2021 foram apresentados 342 trabalhos no evento.



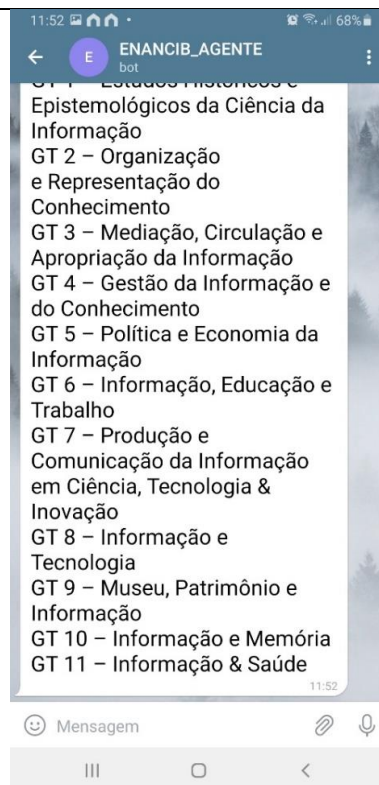
Em 2021 como se dividiu o evento?

Em todos os anos do evento, os trabalhos são divididos em Grupos de Trabalhos (GTs) de acordo com os assuntos abordados e os temas das apresentações.



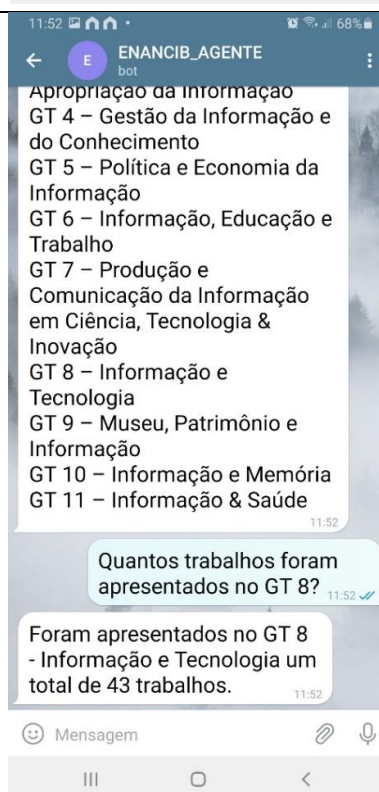
Quais os grupos de trabalhos de 2021?

Em 2021 foram divididos em GTs da seguinte forma:
GT 1 – Estudos Históricos e Epistemológicos da Ciência da Informação
GT 2 – Organização e Representação do Conhecimento
GT 3 – Mediação, Circulação e Apropriação da Informação
GT 4 – Gestão da Informação e do Conhecimento
GT 5 – Política e Economia da Informação
GT 6 – Informação, Educação e Trabalho
GT 7 – Produção e Comunicação da Informação em Ciência, Tecnologia & Inovação
GT 8 – Informação e Tecnologia
GT 9 – Museu, Patrimônio e Informação
GT 10 – Informação e Memória
GT 11 – Informação & Saúde"



Quantos trabalhos foram apresentados no GT 8?

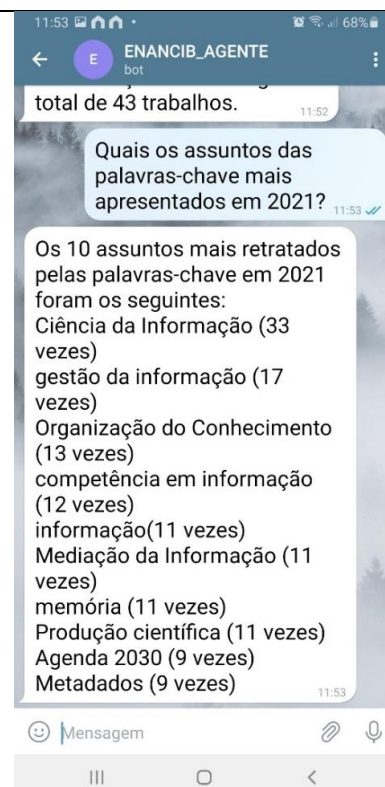
Foram apresentados no GT 8 - Informação e Tecnologia um total de 43 trabalhos.



Quais os assuntos das palavras-chave mais apresentados em 2021?

Os 10 assuntos mais retratados pelas palavras-chave em 2021 foram os seguintes:

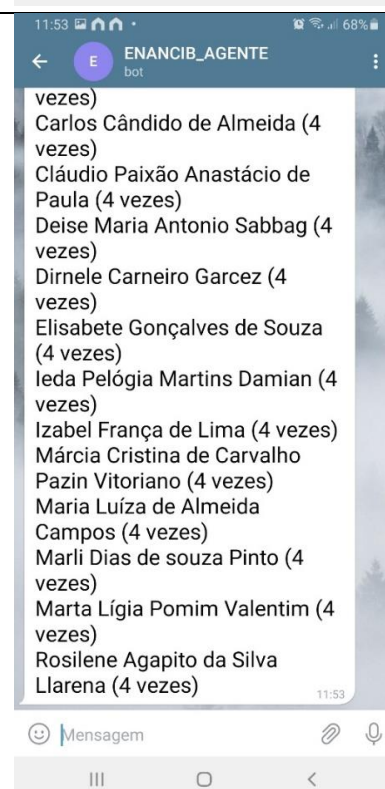
Ciência da Informação (33 vezes)
gestão da informação (17 vezes)
Organização do Conhecimento (13 vezes)
competência em informação (12 vezes)
informação (11 vezes)
Mediação da Informação (11 vezes)
memória (11 vezes)
Produção científica (11 vezes)
Agenda 2030 (9 vezes)
Metadados (9 vezes)"



Quais os autores que mais apresentaram trabalhos em 2021?

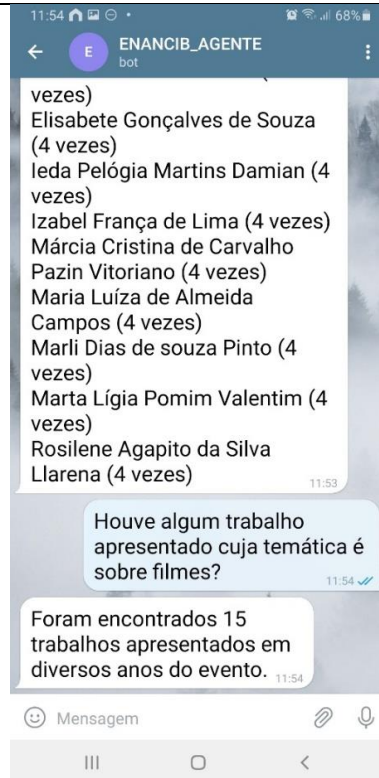
Os autores que mais apresentaram em 2021 foram os seguintes:

Dalton Lopes Martins (5 vezes)
Gustavo Silva Saldanha (5 vezes)
Silvana Aparecida Borsetti Gregório Vidotti (5 vezes)
Alzira Karla Araújo da Silva (4 vezes)
Carlos Cândido de Almeida (4 vezes)
Cláudio Paixão Anastácio de Paula (4 vezes)
Deise Maria Antonio Sabbag (4 vezes)
Dirnele Carneiro Garcez (4 vezes)
Elisabete Gonçalves de Souza (4 vezes)
Ieda Pelógia Martins Damian (4 vezes)
Izabel França de Lima (4 vezes)
Márcia Cristina de Carvalho Pazin Vitoriano (4 vezes)
Maria Luíza de Almeida Campos (4 vezes)
Marli Dias de souza Pinto (4 vezes)
Marta Lúcia Pomim Valentim (4 vezes)
Rosilene Agapito da Silva Llarena (4 vezes)"



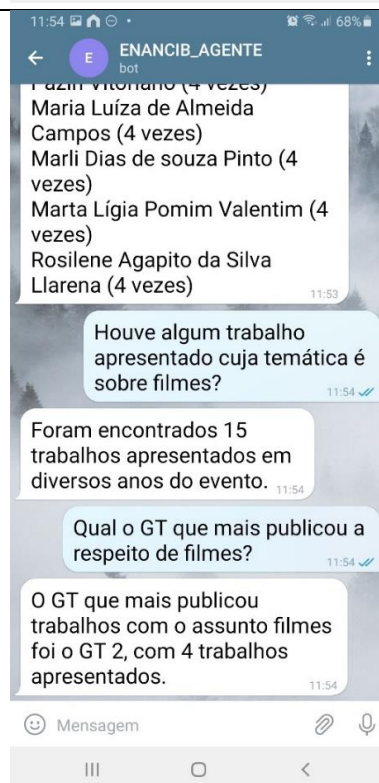
Houve algum trabalho apresentado cuja temática é sobre filmes?

Foram encontrados 15 trabalhos apresentados em diversos anos do evento.



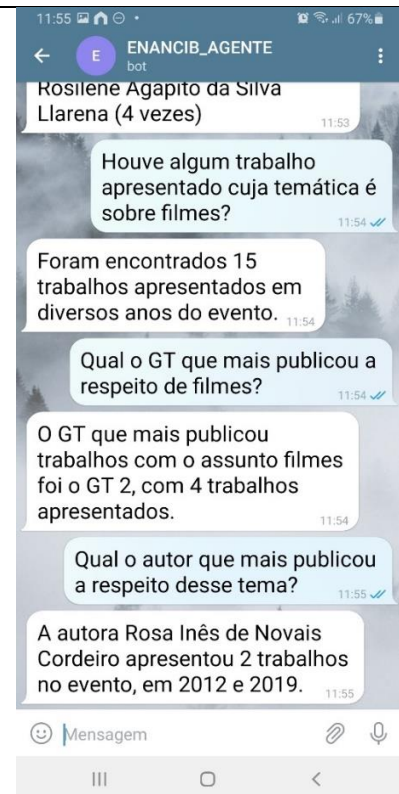
Qual o GT que mais publicou a respeito de filmes?

O GT que mais publicou trabalhos com o assunto filmes foi o GT 2, com 4 trabalhos apresentados.



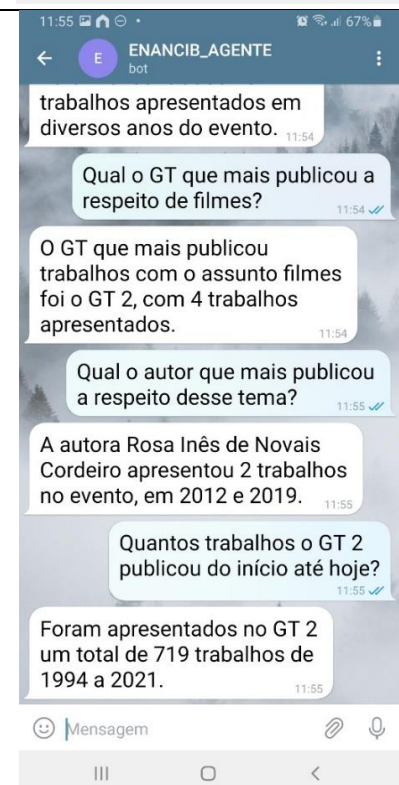
Qual o autor que mais publicou a respeito desse tema?

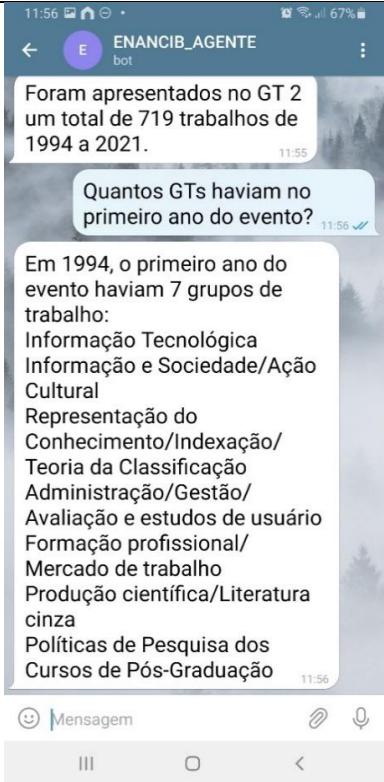
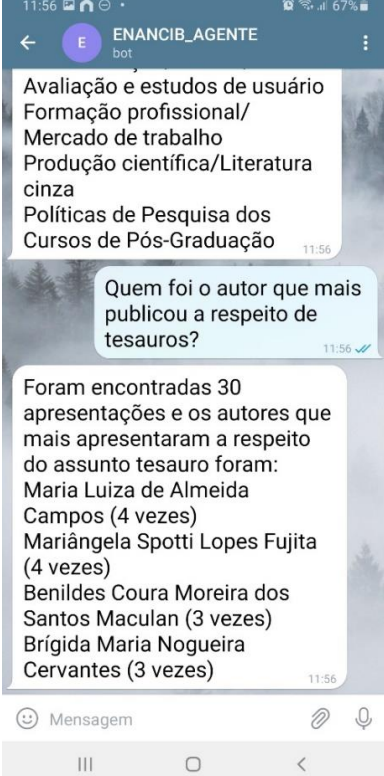
A autora Rosa Inês de Novais Cordeiro apresentou 2 trabalhos no evento, em 2012 e 2019.



Quantos trabalhos o GT 2 publicou do início até hoje?

Foram apresentados no GT 2 um total de 719 trabalhos de 1994 a 2021.

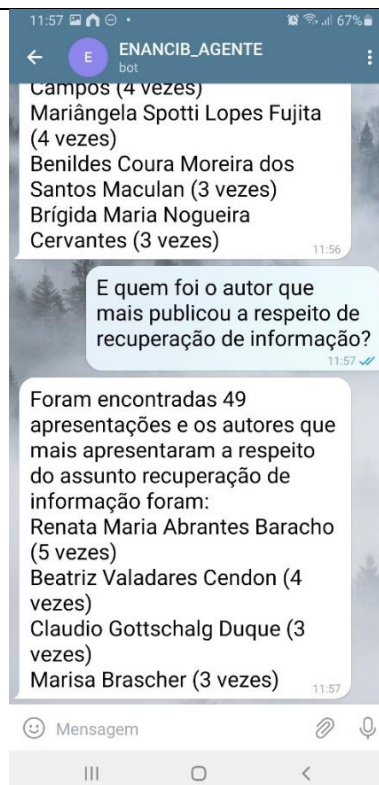


Quantos GTs havia no primeiro ano do evento? Em 1994, o primeiro ano do evento havia 7 grupos de trabalho: Informação Tecnológica Informação e Sociedade/Ação Cultural Representação do Conhecimento/Indexação/Teoria da Classificação Administração/Gestão/Avaliação e estudos de usuário Formação profissional/Mercado de trabalho Produção científica/Literatura cinza Políticas de Pesquisa dos Cursos de Pós-Graduação"	
Quem foi o autor que mais publicou a respeito de tesouros? Foram encontradas 30 apresentações e os autores que mais apresentaram a respeito do assunto tesouro foram: Maria Luiza de Almeida Campos (4 vezes) Mariângela Spotti Lopes Fujita (4 vezes) Benildes Coura Moreira dos Santos Maculan (3 vezes) Brígida Maria Nogueira Cervantes (3 vezes)"	

E quem foi o autor que mais publicou a respeito de recuperação de informação?

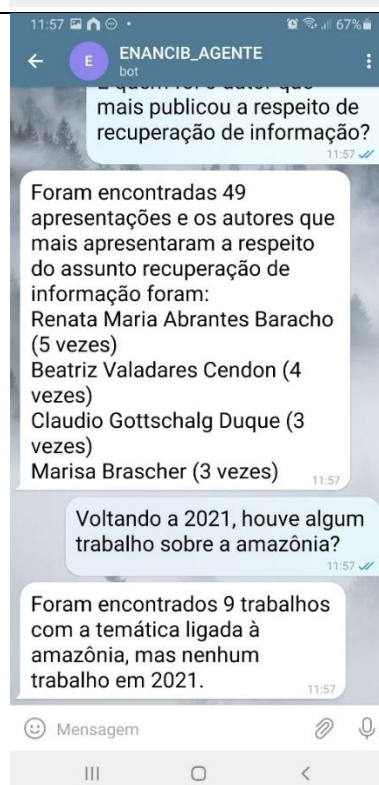
Foram encontradas 49 apresentações e os autores que mais apresentaram a respeito do assunto recuperação de informação foram:

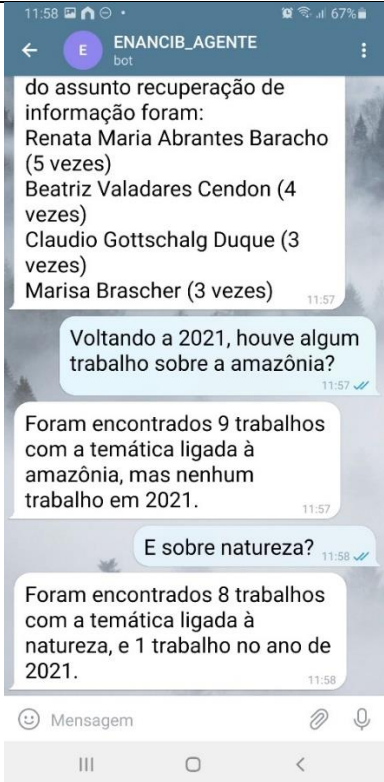
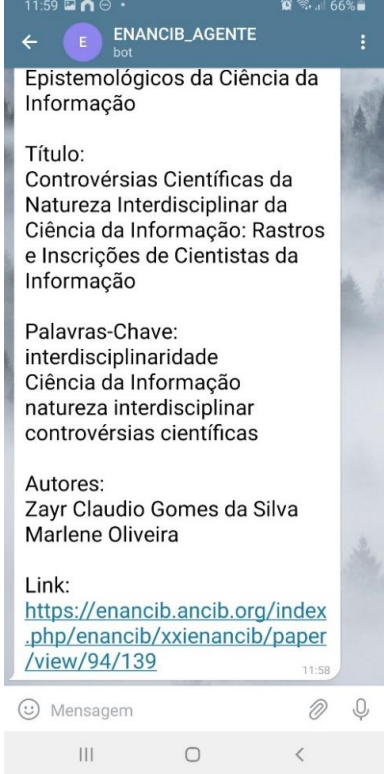
Renata Maria Abrantes Baracho (5 vezes)
Beatriz Valadares Cendon (4 vezes)
Claudio Gottschalg Duque (3 vezes)
Marisa Brascher (3 vezes)"



Voltando a 2021, houve algum trabalho sobre a Amazônia?

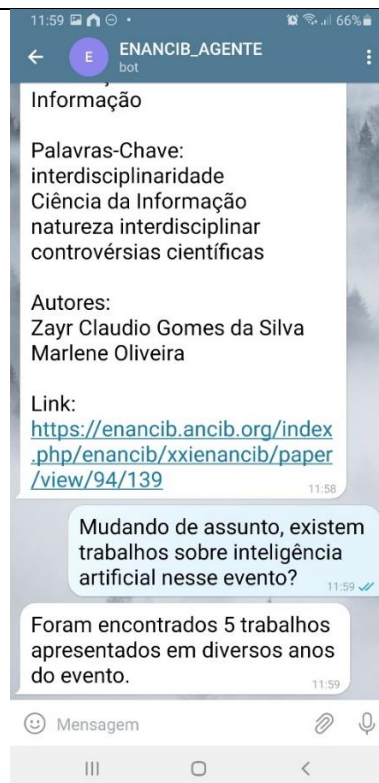
Foram encontrados 9 trabalhos com a temática ligada à Amazônia, mas nenhum trabalho em 2021.



E sobre natureza? Foram encontrados 8 trabalhos com a temática ligada à natureza, e 1 trabalho no ano de 2021.	 do assunto recuperação de informação foram: Renata Maria Abrantes Baracho (5 vezes) Beatriz Valadares Cendon (4 vezes) Claudio Gottschalg Duque (3 vezes) Marisa Brascher (3 vezes) Voltando a 2021, houve algum trabalho sobre a amazônia? Foram encontrados 9 trabalhos com a temática ligada à amazônia, mas nenhum trabalho em 2021. E sobre natureza? Foram encontrados 8 trabalhos com a temática ligada à natureza, e 1 trabalho no ano de 2021.
Me mostre este trabalho. Ano: 2021 GT 1 - Estudos Históricos e Epistemológicos da Ciência da Informação Título: Controvérsias Científicas da Natureza Interdisciplinar da Ciência da Informação: Rastros e Inscrições de Cientistas da Informação Palavras-Chave: interdisciplinaridade Ciência da Informação natureza interdisciplinar controvérsias científicas Autores: Zayr Claudio Gomes da Silva Marlene Oliveira Link: https://enancib.ancib.org/index.php/enancib/xxienancib/paper/view/94/139	 Epistemológicos da Ciência da Informação Título: Controvérsias Científicas da Natureza Interdisciplinar da Ciência da Informação: Rastros e Inscrições de Cientistas da Informação Palavras-Chave: interdisciplinaridade Ciência da Informação natureza interdisciplinar controvérsias científicas Autores: Zayr Claudio Gomes da Silva Marlene Oliveira Link: https://enancib.ancib.org/index.php/enancib/xxienancib/paper/view/94/139

Mudando de assunto, existem trabalhos sobre inteligência artificial nesse evento?

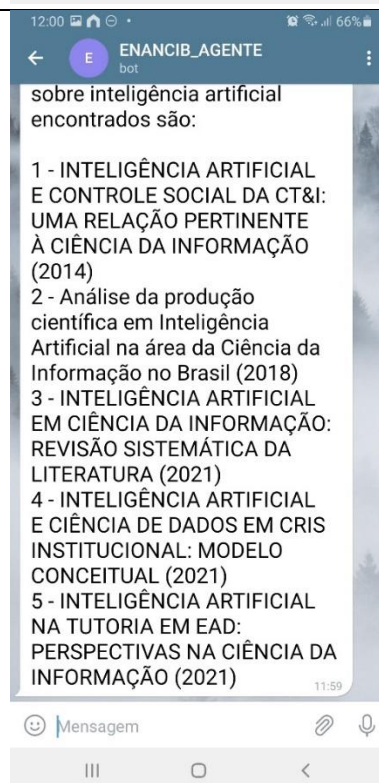
Foram encontrados 5 trabalhos apresentados em diversos anos do evento.



Me mostre os títulos.

Os 5 trabalhos com a temática sobre inteligência artificial encontrados são:

- 1 - INTELIGÊNCIA ARTIFICIAL E CONTROLE SOCIAL DA CT&I: UMA RELAÇÃO PERTINENTE À CIÊNCIA DA INFORMAÇÃO (2014)
- 2 - Análise da produção científica em Inteligência Artificial na área da Ciência da Informação no Brasil (2018)
- 3 - INTELIGÊNCIA ARTIFICIAL EM CIÊNCIA DA INFORMAÇÃO: REVISÃO SISTEMÁTICA DA LITERATURA (2021)
- 4 - INTELIGÊNCIA ARTIFICIAL E CIÊNCIA DE DADOS EM CRIS INSTITUCIONAL: MODELO CONCEITUAL (2021)
- 5 - INTELIGÊNCIA ARTIFICIAL NA TUTORIA EM EAD: PERSPECTIVAS NA CIÊNCIA DA INFORMAÇÃO (2021)"



Me mostre o trabalho 3

Ano: 2021

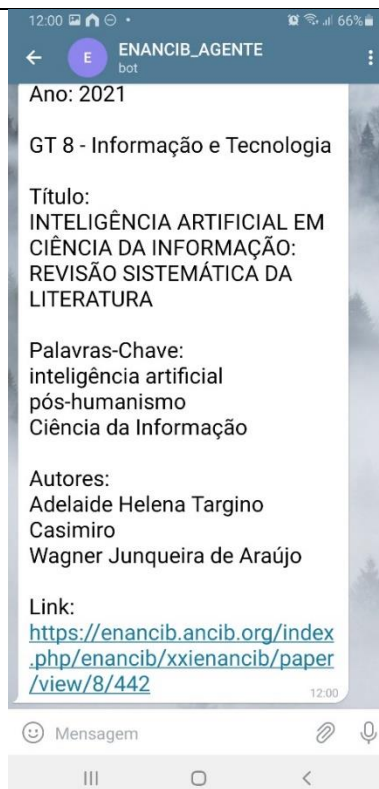
GT 8 - Informação e Tecnologia

Título:
INTELIGÊNCIA ARTIFICIAL EM CIÊNCIA DA
INFORMAÇÃO: REVISÃO SISTEMÁTICA DA
LITERATURA

Palavras-Chave:
inteligência artificial
pós-humanismo
Ciência da Informação

Autores:
Adelaide Helena Targino Casimiro
Wagner Junqueira de Araújo

Link:
<https://enancib.ancib.org/index.php/enancib/xxienancib/paper/view/8/442>



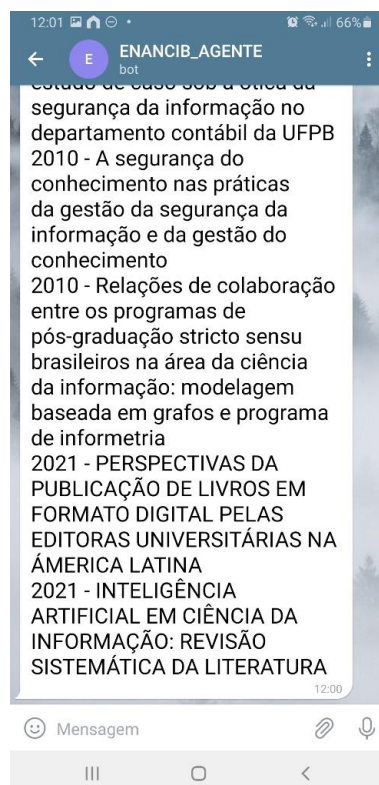
Estes autores já publicaram em outros anos do evento?

Outras publicações destes autores são:

Adelaide Helena Targino Casimiro
2018 - GESTÃO DO CONHECIMENTO E PRÁTICAS
ARQUIVÍSTICAS: CONSTRUTOS TEÓRICOS PARA
FORMAÇÃO E PRESERVAÇÃO DA MEMÓRIA
ORGANIZACIONAL
2021 - MEDIAÇÃO DA INFORMAÇÃO E COMPETÊNCIA
EM INFORMAÇÃO: UMA RELAÇÃO POSSÍVEL EM
CENÁRIO PANDÊMICO
2021 - INTELIGÊNCIA ARTIFICIAL EM CIÊNCIA DA
INFORMAÇÃO: REVISÃO SISTEMÁTICA DA
LITERATURA

Wagner Junqueira de Araújo

2019 - PROSPECÇÃO DE CENÁRIOS: ESTUDO SOBRE A
IMPLANTAÇÃO DO PROCESSO ELETRÔNICO
NACIONAL NA UNIVERSIDADE FEDERAL DA PARAÍBA
2018 - VALIDAÇÃO DO MAPEAMENTO DE FLUXOS DE
INFORMAÇÃO EM PROCESSOS ORGANIZACIONAIS:
UMA ABORDAGEM COM FOCO ARQUIVÍSTICO
2018 - CIBER TERRORISMO NA PARAÍBA
2017 - ESTUDO DE CENÁRIOS PROSPECTIVOS PARA O
MERCADO EDITORIAL DO LIVRO DIGITAL NO BRASIL
2017 - SITES DOS MUNICÍPIOS DA PARAÍBA: ANÁLISE
DE VULNERABILIDADES COMPUTACIONAIS
2016 - ANÁLISE DA DIMENSÃO HUMANA NO
PROCESSO DE GESTÃO DE SEGURANÇA DA
INFORMAÇÃO
2016 - APLICAÇÃO DO OPEN MONOGRAPH PRESS POR
EDITORAS BRASILEIRAS
2015 - ASPECTOS HUMANOS NA GESTÃO DA
SEGURANÇA DA INFORMAÇÃO



2015 - Aspectos de Vulnerabilidades em Bibliotecas Digitais Acessíveis 2014 - ANÁLISE DO SISTEMA DE INFORMAÇÃO DA BIBLIOTECA CENTRAL DA UFPB NOS PROCESSOS DE GESTÃO DA INFORMAÇÃO PARA O SETOR DE REFERÊNCIA 2013 - ANÁLISE DE INFORMAÇÕES PESSOAIS NA WEB: métrica para identificar o grau de exposição da informação 2013 - ANÁLISE DE RISCO NO SISTEMA DE CONCESSÃO DE DIÁRIAS E PASSAGENS (SCDP): estudo de caso sob a ótica da segurança da informação no departamento contábil da UFPB 2010 - A segurança do conhecimento nas práticas da gestão da segurança da informação e da gestão do conhecimento 2010 - Relações de colaboração entre os programas de pós-graduação stricto sensu brasileiros na área da ciência da informação: modelagem baseada em grafos e programa de informetria 2021 - PERSPECTIVAS DA PUBLICAÇÃO DE LIVROS EM FORMATO DIGITAL PELAS EDITORAS UNIVERSITÁRIAS NA ÁMERICA LATINA 2021 - INTELIGÊNCIA ARTIFICIAL EM CIÊNCIA DA INFORMAÇÃO: REVISÃO SISTEMÁTICA DA LITERATURA"	
Obrigado pela ajuda, Bye Bye Agradeço a sua visita e espero que eu tenha ajudado. Sempre que precisar estarei aqui à disposição. Até mais e tenha uma Boa Semana.	