

Análise de Artigos Científicos sobre Internet das Coisas (IoT) com PLN

Gustavo Henrique Trassi Ganaza

João Augusto da Silva Gomes

Yoshiyuki Fugie

16 de julho de 2025

Universidade Estadual de Maringá

Curso de Ciência da Computação

Disciplina: Introdução a Inteligência Artificial

Professor Dr. Wagner Igarashi

Introdução

- **Objetivo:** Analisar um conjunto de 10 artigos científicos da área de Internet das Coisas (IoT), extraídos do portal IEEE Xplore, utilizando técnicas de Processamento de Linguagem Natural (PLN).
- **Escopo:**
 - Leitura e pré-processamento de textos em formato PDF.
 - Extração de informações estruturadas como objetivo, problema, método e contribuição de cada artigo.
 - Identificação de tendências, termos-chave, técnicas mencionadas e direções para trabalhos futuros por meio de análises de frequência e visualização de dados.

- **Processamento de Linguagem Natural (PLN):** Campo da inteligência artificial focado na interação entre computadores e a linguagem humana.
- **Pré-processamento de Texto:**
 - **Lematização (spaCy):** Redução das palavras à sua forma base (lema) para agrupar termos com o mesmo significado.
 - **Remoção de Stopwords (NLTK):** Exclusão de palavras comuns (ex: "o", "que", "de") que não agregam valor semântico à análise.

- **Modelo Bag-of-Words:** Utilizado para representar os textos como um conjunto de palavras, permitindo a análise de frequência (unigramas) para identificar os termos mais relevantes.
- **Extração de Informação:** Uso de expressões regulares (regex) para buscar e extrair sentenças-chave que descrevem o objetivo, problema, método e contribuição dos artigos.

Materiais e Métodos

- **Artigos Científicos:** 10 artigos selecionados do portal IEEE Xplore, focando em temas relevantes de IoT.
- **Ferramentas e Bibliotecas:**
 - Python: Linguagem de programação.
 - PyPDF2 e spaCy: Para extração de texto limpo e pré-processamento (lematização, tokenização).
 - NLTK: Para remoção de stopwords.
 - Pandas: Para manipulação e organização dos dados extraídos.
 - WordCloud e Matplotlib: Para a geração das visualizações de dados (nuvem de palavras, gráficos).

Título do Artigo	Ano
A simplified approach for dynamic security risk management in connected and autonomous vehicles	2019
Automation in Agriculture and IoT	2019
Cloud for IoT - A survey of technologies and security features of public cloud IoT solutions	2019
Developing complex services in an IoT ecosystem	2015
Emerging privacy challenges and approaches in CAV systems	2019
IoT Challenges and Issues in Indian Perspective	2018
Next Generation Lightweight Cryptography for Smart IoT Devices Implementation Challenges and Applications	2019
Secure - zero touch device onboarding	2019
Towards an integrated privacy protection framework for IoT Contextualising regulatory requirements with industry best practices	2019
Towards an understanding of emerging cyber security threats in mapping the IoT	2019

1. Leitura e Limpeza: Extração de texto dos PDFs, removendo cabeçalhos e rodapés recorrentes.
2. Pré-processamento: Aplicação de lematização e remoção de stopwords.
3. Análise de Frequência: Contagem de unigramas para identificar os termos mais citados.
4. Extração de Informações: Aplicação de padrões (regex) para identificar as seções de interesse.
5. Visualização: Geração de gráficos e nuvens de palavras para ilustrar os resultados.

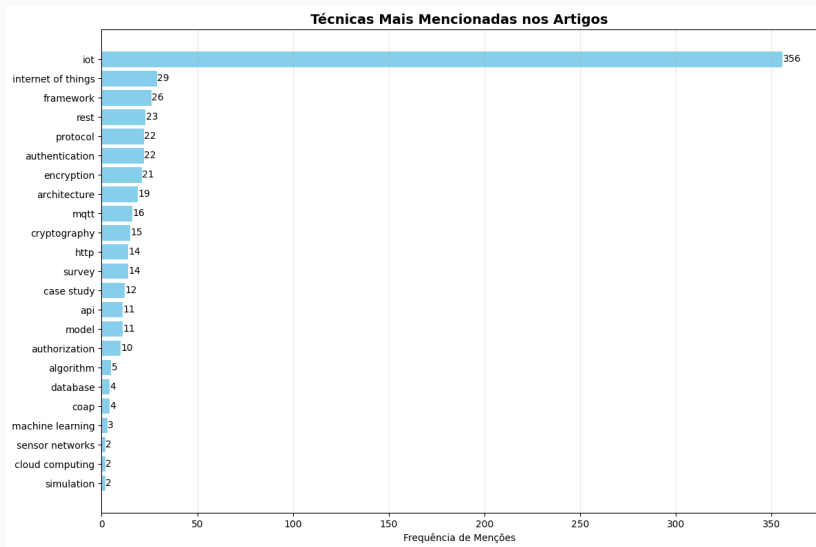
Resultados



- A nuvem de palavras destaca os termos mais frequentes em todos os artigos, fornecendo uma visão geral dos principais focos de pesquisa.
- Termos como "iot", "privacy", "data", "device" e "security" são os mais proeminentes, indicando que a segurança e a privacidade de dados em dispositivos IoT são os temas centrais dos artigos analisado.

Termo	Frequência
iot	350
privacy	248
datum	237
device	219
service	216
use	173
security	159
system	157
risk	129
attack	121

Técnicas e Métodos Mencionados



Técnicas e Métodos Mencionados

- A análise das técnicas revela um forte foco em aspectos de comunicação e segurança.
- Termos como "iot", "authentication", "rest" (protocolo de API), "mqtt" (protocolo de mensagens) e "encryption" aparecem com alta frequência.
- Isso sugere que os artigos tratam primariamente da arquitetura e da proteção de ecossistemas IoT, em vez de focar em técnicas de análise de dados como "machine learning" ou "deep learning", que tiveram menções menos expressivas no conjunto de artigos.

Técnicas e Métodos Mencionados por Artigo

Developing_complex_services_in_an_IoT_ecosystem.pdf:

iot: 26

internet of things: 3

architecture: 3

rest: 3

http: 3

Emerging privacy challenges and approaches in CAV systems.pdf:

cryptography: 3

authentication: 3

internet of things: 3

protocol: 3

iot: 2

IoT_Challenges_and_Issues_in_Indian_Perspective.pdf:

iot: 48

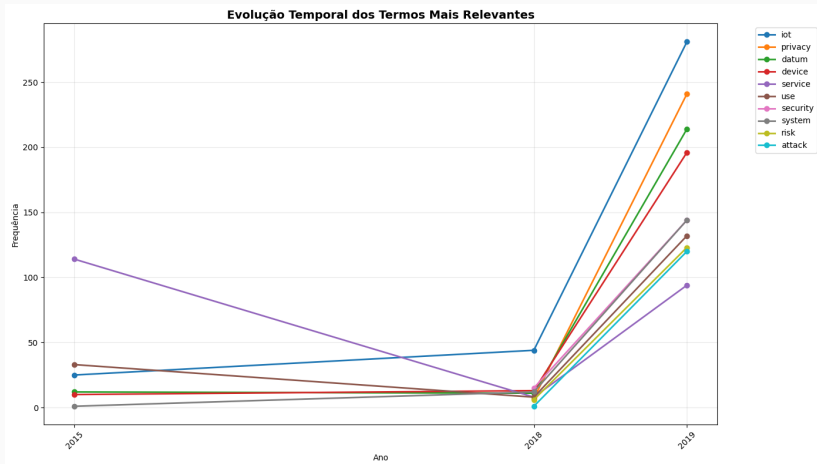
internet of things: 7

survey: 5

authentication: 1

protocol: 1

Evolução Temporal dos Termos



Evolução Temporal dos Termos

- O gráfico ilustra a frequência dos termos mais relevantes ao longo dos anos dos artigos publicados (2015, 2018, 2019).
- Nota-se uma alta concentração de artigos em 2019, ano em que termos como "iot", "privacy" e "data" atingiram seu pico de menções no conjunto de dados.
- No entanto, isso se deve ao fato de que a maioria dos artigos analisados é de 2019, o que pode distorcer a percepção da evolução temporal.

Tópicos para Trabalhos Futuros



Figura 1: Nuvem de palavras que foram identificadas como tópicos para trabalhos futuros nos artigos analisados.

Tópicos para Trabalhos Futuros

- Foram encontrados Trabalhos Futuros em 8 artigos.
- Nessses 8 artigos, foram identificadas 14 sentenças que descrevem possíveis direções para pesquisas futuras.
- Analisando essas sentenças, percebe-se que as pesquisas apontam para a continuidade no desenvolvimento de sistemas, segurança e abordagens baseadas em conhecimento.

Exemplo de Sentenças Extraídas para Trabalhos Futuros

- "We anticipate that some of our future work will be to develop more uses of the knowledge-based system on the security analysis, and to extend this research in simulation environment for further ve..."
- "This work categorises the emerging privacy challenges and solutions for CAV systems and identifies the knowledge gap for future research, which will minimise and mitigate privacy concerns without ham..."

Conclusão

- A análise dos artigos científicos sobre IoT revelou um foco significativo em segurança, privacidade e comunicação entre dispositivos.
- As técnicas de Processamento de Linguagem Natural (PLN) foram eficazes para extrair informações relevantes e identificar tendências na literatura.
- A evolução temporal dos termos sugere um aumento do interesse em temas relacionados à segurança e privacidade nos últimos anos.
- Os tópicos para trabalhos futuros indicam uma continuidade nas pesquisas, com ênfase em segurança, desenvolvimento de sistemas e abordagens baseadas em conhecimento.

Referências

- **spaCy.** Industrial-Strength Natural Language Processing. Disponível em: <https://spacy.io/>
- **NLTK Project.** Natural Language Toolkit. Disponível em: <https://www.nltk.org/>
- **PyPDF2.** A pure-python PDF library. Disponível em: <https://pypdf2.readthedocs.io/>
- **WordCloud.** A little word cloud generator in Python. Disponível em: https://github.com/amueller/word_cloud
- **Python re module.** Regular expression operations. Disponível em: <https://docs.python.org/3/library/re.html>
- **IEEE Xplore Digital Library.** Disponível em: <https://ieeexplore.ieee.org/>