

Construção de uma ontologia baseada na DBLP

*Construction of an ontology based on DBLP*

Vitor Duarte Bezerra de OLIVEIRA<sup>1</sup>

---

<sup>1</sup> Universidade Federal Rural do Semi-Árido, Departamento de computação, Ciência da computação. R. Francisco Mota, 572 – Bairro Costa e Silva, 59.625-900, Mossoró, RN, Brasil. Correspondência para/*Correspondence to*: Vitor Duarte Bezerra de OLIVEIRA. *E-mail*: vitor.oliveira@alunos.ufersa.edu.br

## Resumo

Este artigo aborda a criação de uma ontologia voltada para a base de dados de publicações acadêmicas DBLP (*Digital Bibliography & Library Project*). O objetivo principal é demonstrar como as modernas ferramentas de modelagem e desenvolvimento de ontologias, como o *Visual Paradigm*, *OntoUML* e *Protégé*, podem ser eficazmente empregadas para construir uma estrutura de conhecimento que facilite a pesquisa, navegação e recuperação de informações em um domínio tão vasto. A construção da ontologia envolveu a seleção de conceitos, propriedades e relações relevantes para o domínio acadêmico (DBLP). O *Visual Paradigm* foi utilizado como ferramenta de modelagem, permitindo a representação visual de conceitos e relações complexas. O uso da *OntoUML* como linguagem de modelagem ofereceu uma maneira precisa de descrever e relacionar os elementos da ontologia, garantindo uma representação coerente e estruturada. A metodologia detalhada neste artigo inclui a modelagem conceitual dos principais elementos da DBLP, a representação desses elementos por meio de *OntoUML* e a implementação da ontologia usando o ambiente *Protégé*. Os resultados e conclusões evidenciam como a ontologia da DBLP pode ser aplicada na prática, simplificando a busca, a análise e a navegação nas vastas informações disponíveis na base de dados.

**Palavras-chave:** Ontologia. DBLP. *Visual Paradigm*. *Onto UML*. *Protégé*.

## Abstract

*This article addresses the creation of an ontology focused on the academic publications database DBLP (Digital Bibliography & Library Project). The main objective is to demonstrate how modern ontology modeling and development tools, such as Visual Paradigm, OntoUML, and Protégé, can be effectively employed to build a knowledge structure that facilitates research, navigation, and information retrieval in a domain as vast. The construction of the ontology involved a selection of concepts, properties, and relationships relevant to the academic domain (DBLP). Visual Paradigm was used as a modeling tool, enabling the visual representation of complex concepts and relationships. The use of OntoUML as a modeling language provided a precise way to describe and relate ontology elements, ensuring a coherent and structured representation. The methodology detailed in this article includes the conceptual modeling of key elements within DBLP, representing these elements through OntoUML, and implementing the ontology using the Protégé environment. The results and conclusions highlight how the DBLP ontology can be practically applied, simplifying the search, analysis, and navigation of the vast information available in the database.*

**Keywords:** Ontology. DBLP. *Visual Paradigm*. *OntoUML*. *Protégé*.

## Introdução

“Uma ontologia é um modelo explícito de um domínio, que compreende os conceitos, propriedades e relações relevantes para esse domínio. Ela atua como um arcabouço para representar o conhecimento de um domínio de forma estruturada e organizada.” (Gruber, T. R, 1993, p.5). A criação de ontologias desempenha um papel importantíssimo na organização e na representação do conhecimento em domínios complexos. Além disso, o uso de ontologias pode facilitar a integração de dados de diferentes fontes, a conexão entre sistemas e a análise avançada de dados, contribuindo para o avanço da pesquisa e do conhecimento em diversas áreas.

Neste contexto, este presente descreve a criação de uma ontologia específica para a base de dados de publicações acadêmicas DBLP (*Digital Bibliography & Library Project*), fazendo uso de ferramentas como o *Visual Paradigm* em conjunto com *OntoUML*, para a modelagem da ontologia e posteriormente do *Protégé* para os passos finais de desenvolvimento da ontologia.

A DBLP é um repositório abrangente de trabalhos acadêmicos na área da ciência da computação e abrange uma ampla gama de tópicos, autores e conferências. O *Visual Paradigm* é uma ferramenta de modelagem que pode ser utilizada para criar ontologias. Ela oferece uma variedade de recursos para facilitar a criação e a manutenção de ontologias quando usada em conjunto com *OntoUML*. “*OntoUML* é uma extensão da linguagem de modelagem UML para a representação de ontologias” (Fensel et al., *Ontology languages for knowledge management*, 2009). Ela oferece uma variedade de recursos para facilitar a representação de ontologias em UML, incluindo: classes ontológicas (são usadas para representar conceitos em uma ontologia), propriedades ontológicas (usadas para representar propriedades de conceitos em uma ontologia) e as relações ontológicas (usadas para representar relações entre conceitos em uma ontologia).

O *Protégé*, por sua vez, é um ambiente de desenvolvimento de ontologias que simplifica a criação, edição e gerenciamento dessas estruturas de conhecimento. Com suas capacidades de raciocínio e inferência, o *Protégé* transforma a ontologia em uma ferramenta dinâmica que vai além da simples representação estática.

A criação da ontologia direcionada à DBLP, por meio do uso dessas ferramentas, assume uma relevância significativa na promoção da pesquisa acadêmica e na aprimoração da acessibilidade às informações relacionadas à ciência da computação.

## Procedimentos Metodológicos

A criação da ontologia como o tema DBLP envolveu uma série de procedimentos metodológicos. Estas etapas foram essenciais para garantir a representação precisa dos conceitos, propriedades e relações no domínio acadêmico diversificado da DBLP. A primeira fase da metodologia envolveu a seleção cuidadosa de conceitos, propriedades e relações

relevantes para o domínio da DBLP e a construção do documento de requisitos da ontologia (ORSD). Essa etapa foi conduzida com base em uma revisão da literatura e análise das necessidades. Os conceitos-chave, como publicações e autores foram identificados e definidos.

Após a definição dos conceitos, a modelagem conceitual foi realizada usando a ferramenta *Visual Paradigm*. Essa etapa permitiu a representação visual clara dos conceitos e das relações entre eles. O *Visual Paradigm* entrega uma interface simples de modelagem de classes e recursos avançados de modelagem, tornando o processo mais eficiente e compreensível (figuras 1, 2 e 3).

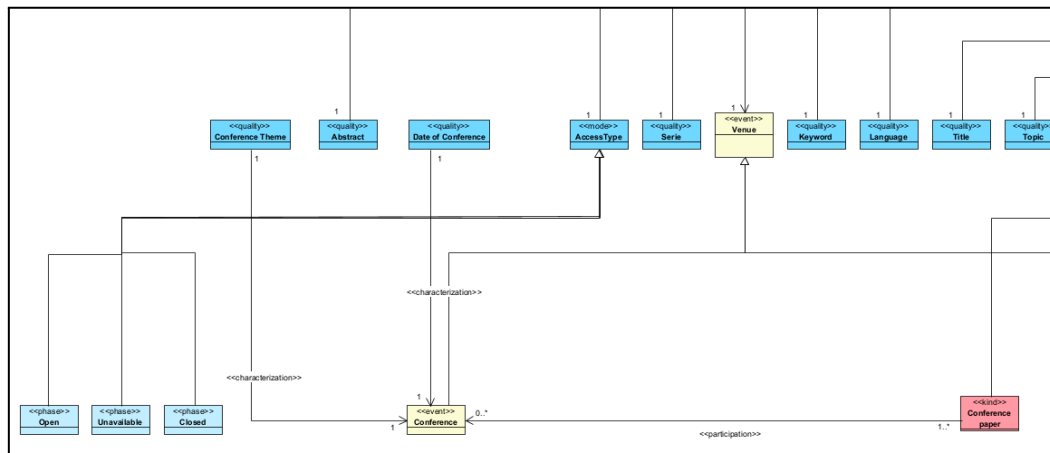


Figura 1. Visão mais à esquerda

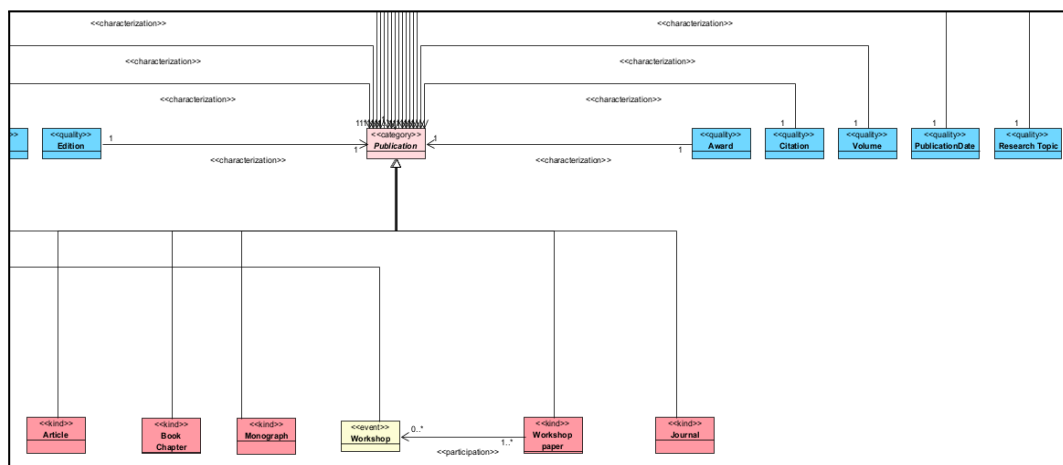


Figura 2. Visão central

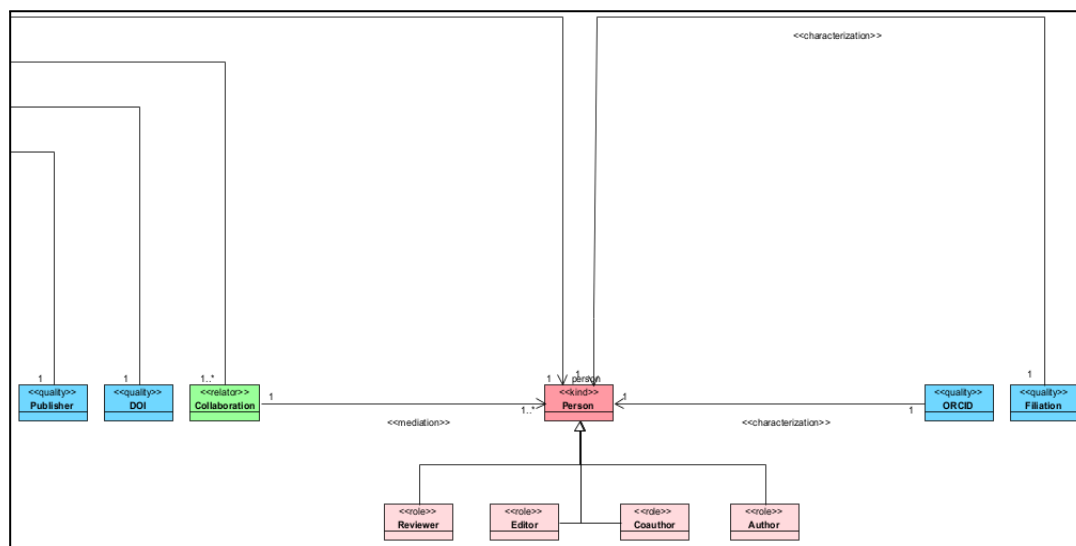


Figura 3. Visão mais à direita

Com a modelagem conceitual pronta, a linguagem *OntoUML* foi adotada para uma representação detalhada dos conceitos e relações. Ela oferece recursos específicos para ontologias, incluindo classes ontológicas, propriedades ontológicas e relações ontológicas, o que permitiu uma descrição precisa e estruturada dos elementos da ontologia.

O próximo passo envolveu a implementação da ontologia usando o ambiente *Protégé*. O *Protégé* é um ambiente de desenvolvimento de ontologias que oferece recursos avançados de raciocínio e inferência. Isso transformou a ontologia em uma ferramenta dinâmica que pode auxiliar na pesquisa, análise e navegação de informações na DBLP.

Com o modelo pronto, após a fase do *Protégé*, foi o momento de responder as questões que foram estabelecidas no documento de requisitos da ontologia (figura 4). Para isso fez-se o uso de consultas SPARQL (SPARQL Protocol and RDF Query Language). O SPARQL é uma linguagem de consulta amplamente reconhecida para dados ligados e ontologias RDF. Ele permite a recuperação de informações específicas da ontologia e a realização de análises avançadas.

#### b. Functional Requirements: Groups of Competency Questions

CQG1. Comparing authors and academic production at UFERSA (Federal Rural University of the Semi-Arid) (5 CQ)

CQ1. What is the total number of academic publications (papers, conferences, monographs, etc.) for each year (2001 -2023)?

CQ2. What is the total number of publications per person?

CQ3. What is the quantity of each type of publication?

CQ4. Who published the most in journals?

CQ5. Who published the most in a specific period of time?

CQ6. What were the collaborations between authors and their respective publications?

CQ7. What were the publications during a specific period of time?

Figura 4. Questões de competência definidas no ORSD

Por último, após a conclusão da ontologia, fez-se uso da ferramenta *WebVOWL*, uma poderosa solução de visualização que possibilita a representação interativa e clara das ontologias. Essa ferramenta proporciona uma representação gráfica dos conceitos, propriedades e relações presentes na ontologia, tornando mais acessível aos usuários a compreensão da estrutura e das interconexões entre os elementos. Essa capacidade aprimorada de visualização desempenha um papel fundamental na facilitação da análise da ontologia.

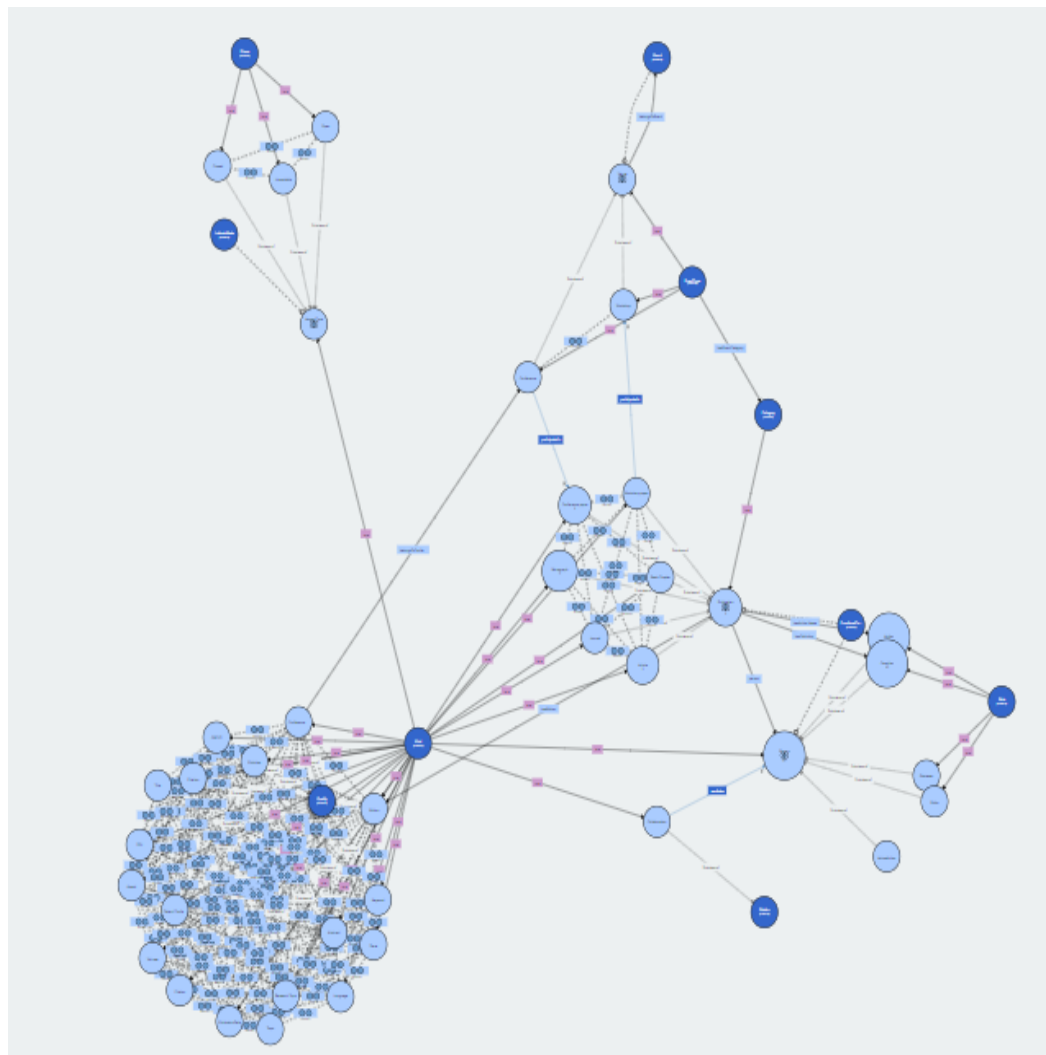


Figura 5. Visualização usando *WebVOWL* com *class disjointness*

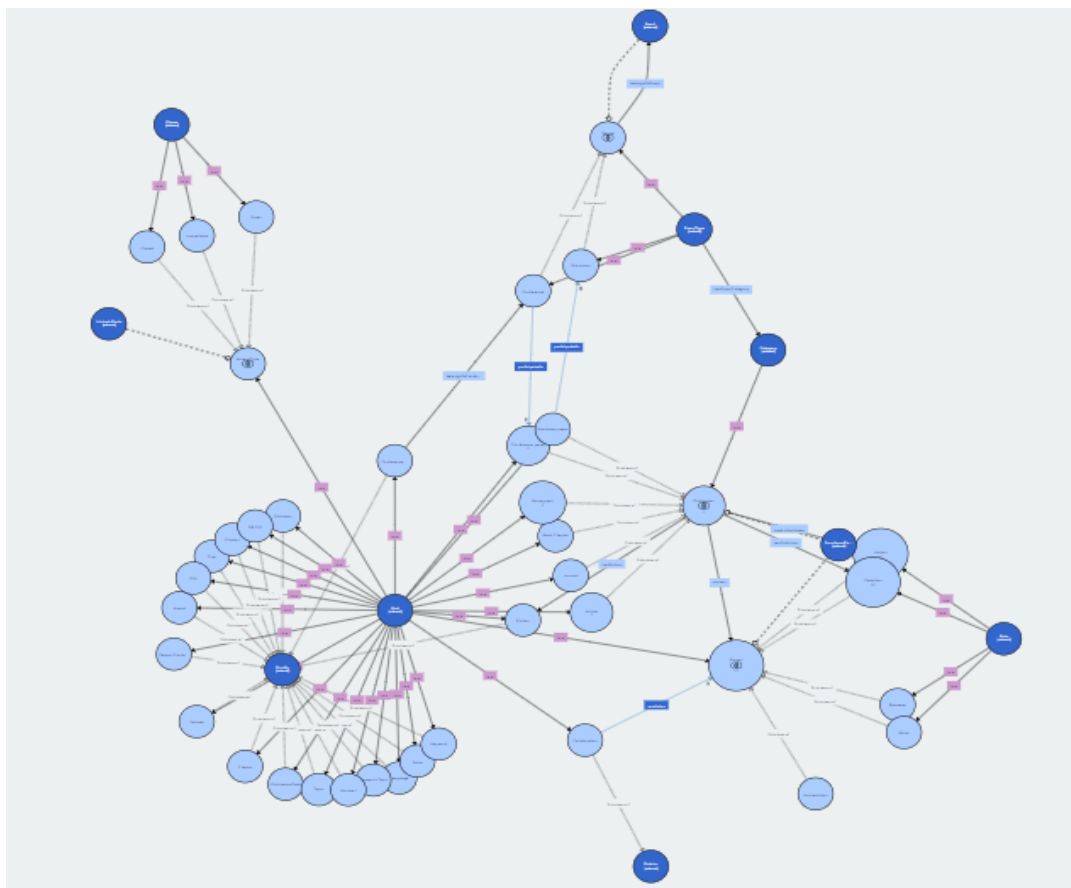


Figura 6. Visualização usando *WebVOWL* sem *class disjointness*

## Resultados

Após a conclusão das etapas da metodologia, a ontologia está finalmente pronta para ser explorada. Os resultados foram significativos e abrem perspectivas promissoras para a pesquisa acadêmica no domínio da DBLP (ciência da computação). Como já mencionado, logo após a modelagem via *OntoUML* e *Visual Paradigm* o modelo de ontologia foi importado para o *Protégé* e ficou com a seguinte estrutura:

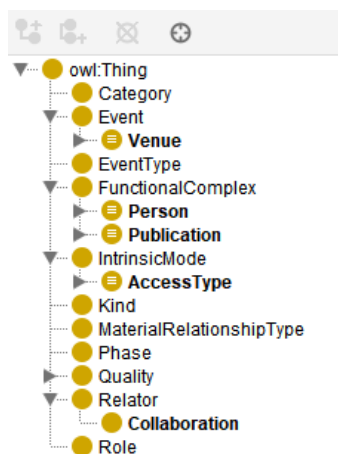


Figura 7. Classes criadas no *Protégé*

Com o modelo importado ao *Protégé*, e após a finalização dos ajustes finais, como criação de instâncias, criação de *object properties* e *data properties*, chega o momento de realizar as consultas SPARQL para obter as respostas das perguntas que foram definidas no ORSD. As consultas serão exibidas com seu devido código (SPARQL) em conjunto com as respostas obtidas.

1. Qual a quantidade de publicações em cada ano (2001 - 2023)?

```
Unset
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX my: <http://example.com#>

SELECT ?year (COUNT(?publication) AS ?publicationCount)
WHERE {
  {
    ?publication my:hasJournalPublicationYear ?year .
  }
  UNION
  {
    ?publication my:hasWorkshopYear ?year .
  }
  UNION
  {
    ?publication my:hasConferenceYear ?year .
  }
}
GROUP BY ?year
ORDER BY ?year
```

?year	?publicationCount
2015	8
2016	6
2017	3
2018	6
2019	7
2020	13
2021	3
2022	12
2023	7

Figura 8. Resposta da pergunta 1

2. Qual a quantidade de publicações de cada indivíduo?

```
Unset
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
```



```

PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX my: <http://example.com#>

SELECT ?person (COUNT(?publication) AS ?publicationCount)
WHERE {
  ?person rdf:type my:Person .
  ?publication rdf:type my:Publication .
  {
    ?publication my:hasAuthorName ?person .
  }
  UNION
  {
    ?publication my:hasCoAuthor ?person .
  }
}
GROUP BY ?person
ORDER BY DESC(?publicationCount)

```

?person	?publicationCount
my:Francisco_Milton_Mendes_Neto	35
my:Patrício_de_Alencar_Silva	31
my:Lenardo_C_E_Silva	17
my:Leiva_Casemiro_Oliveira	11
my:Angélica_Félix_de_Castro	10
my:Bruno_de_Sousa_Monteiro	9
my:Silvio_Roberto_Fernandes	7
my:Paulo_Gabriel_Gadelha_Queiroz	4
my:Yáskara_Y_M_P_Fernandes	3

Figura 9. Resposta da pergunta 2

### 3. Qual a quantidade de cada tipo de publicação?

```

Unset
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX my: <http://example.com#>

SELECT ?publicationType (COUNT(?publication) AS ?publicationCount)
WHERE {
  ?publication rdf:type my:Publication .
  ?publication a ?publicationType .
  FILTER (
    ?publicationType = my:Journal ||
    ?publicationType = my:ConferencePaper ||
    ?publicationType = my:WorkshopPaper ||
    ?publicationType = my:Article ||
    ?publicationType = my:Monograph ||
    ?publicationType = my:BookChapter
  )
}
GROUP BY ?publicationType

```

ORDER BY DESC(?publicationCount)

?publicationType	?publicationCount
my:ConferencePaper	81
my:Journal	20
my:WorkshopPaper	13
my:Monograph	3

Figura 10. Resposta da pergunta 3

#### 4. Qual pessoa publicou mais em um tipo específico (*Journal*)?

Unset

```
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX my: <http://example.com#>
```

```
SELECT ?person (COUNT(?publication) AS ?journalCount)
WHERE {
  ?person rdf:type my:Person .
  ?publication rdf:type my:Publication .
  ?publication a my:Journal .
  {
    ?publication my:hasAuthorName ?person .
  }
  UNION
  {
    ?publication my:hasCoAuthor ?person .
  }
}
GROUP BY ?person
ORDER BY DESC(?journalCount)
```

?person	?journalCount
my:Lenardo_C_E_Silva	6
my:Francisco_Milton_Mendes_Neto	6
my:Leiva_Casemiro_Oliveira	3
my:Fábio_Francisco_da_Costa_Fontes	2
my:Bruno_de_Sousa_Monteiro	1
my:Danniel_Cavalcante_Lopes	1
my:Angélica_Félix_de_Castro	1
my:Patrício_de_Alencar_Silva	1
my:Silvio_Roberto_Fernandes	1

Figura 11. Resposta da pergunta 4

#### 5. Quais foram as Colaborações entre autores?

Unset

```
PREFIX owl: <http://www.w3.org/2002/07/owl#>
```

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX my: <http://example.com#>
```

```
SELECT ?person1 ?person2 ?publication
WHERE {
  ?person1 rdf:type my:Person .
  ?person2 rdf:type my:Person .
  ?publication rdf:type my:Publication .

  {
    ?publication my:hasAuthorName ?person1 .
    ?publication my:hasAuthorName ?person2 .
  }
  UNION
  {
    ?publication my:hasCoAuthor ?person1 .
    ?publication my:hasCoAuthor ?person2 .
  }

  FILTER (?person1 != ?person2)
}
```

?person1	?person2	?publication
my:Lenardo_C_E_Silva	my:Francisco_Milton_Mendes_Neto	my:A_study_on_the_impact_of_moodle_chat_and_forum_interaction_tools_on...
my:Francisco_Milton_Mendes_Neto	my:Lenardo_C_E_Silva	my:A_study_on_the_impact_of_moodle_chat_and_forum_interaction_tools_on...
my:Silvio_Roberto_Fernandes	my:Leiva_Casemiro_Oliveira	my:MOSS_based_Methodology_to_Building_an_Operational_System_for_MIPS...
my:Leiva_Casemiro_Oliveira	my:Silvio_Roberto_Fernandes	my:MOSS_based_Methodology_to_Building_an_Operational_System_for_MIPS...
my:Patricio_de_Alencar_Silva	my:Francisco_Milton_Mendes_Neto	my:An_Investigation_on_the_use_of_Ontologies_for_pattern_classification_...
my:Francisco_Milton_Mendes_Neto	my:Patricio_de_Alencar_Silva	my:An_Investigation_on_the_use_of_Ontologies_for_pattern_classification_...
my:Danniel_Cavalcante_Lopes	my:Angélica_Félix_de_Castro	my:Meteorological_Data_Mining_a_Case_Study_in_the_Brazil_Semi-Arid
my:Angélica_Félix_de_Castro	my:Danniel_Cavalcante_Lopes	my:Meteorological_Data_Mining_a_Case_Study_in_the_Brazil_Semi-Arid
my:Danniel_Cavalcante_Lopes	my:Angélica_Félix_de_Castro	my:Definition_of_Strategies_for_Crime_Prevention_and_Combat_Using_Fuzzy...

36 results

Figura 12. Resposta da pergunta 5

## Discussão

Primeiramente, vale ressaltar que a ontologia, juntamente com o modelo *OntoUML* e as consultas SPARQL estão disponíveis para todos no seguinte repositório: <https://github.com/vitordbo/Web-Ontology-DBLP>.

Tendo em vista que a eficiência na recuperação de informações é um dos principais indicadores de sucesso de uma ontologia semântica. Os resultados positivos obtidos nas consultas SPARQL sugerem que a ontologia está bem projetada e é capaz de recuperar informações de forma precisa e rápida.

Além disso, os resultados destacam o potencial da ontologia para pesquisa acadêmica, permitindo a análise mais aprofundada dos dados da DBLP. O que está em

conformidade com a missão das ontologias semânticas, que é estabelecer uma base robusta para a exploração do conhecimento em domínios específicos por parte de pesquisadores e acadêmicos, conforme destacado por Gangemi et al. (2002).

### **Conclusão**

Em resumo, este artigo oferece uma sólida evidência de que a criação de uma ontologia semântica bem é uma abordagem de grande valor para representar e explorar o conhecimento no contexto da DBLP na ciência da computação e os resultados obtidos através das consultas SPARQL confirmam a eficácia da ontologia.

Levando tudo em consideração, a ontologia desenvolvida não é apenas uma ferramenta útil para pesquisadores, mas também uma contribuição valiosa para o domínio da ciência da computação. Ela proporciona uma base sólida para a organização e exploração de dados na DBLP, o que pode acelerar a descoberta de novas tendências, facilitar a análise de dados e até mesmo promover o avanço do conhecimento na área.

Portanto, a criação dessa ontologia semântica bem-sucedida não apenas fortalece o cenário da ciência da computação, mas também reforça a ideia de que as ontologias são ferramentas poderosas para representar, compartilhar e explorar conhecimento, e que seu uso pode levar a avanços significativos em vários aspectos da sociedade.

### **Referências**

- Gangemi, A., Guarino, N., Masolo, C., Oltramari, A., & Schneider, L. Sweetening Ontologies with DOLCE. In EKAW, p. 166-181, 2002
- Gómez-Pérez, A., Fernández-López, M., & Corcho, O. Ontological Engineering: With Examples from the Areas of Knowledge Management, E-Commerce and the Semantic Web. Springer, 2004.
- Gruber, T. R. A translation approach to portable ontology specifications. *Knowledge acquisition*, v 5(2), p. 3-25, 1993.
- Fensel, D., van Harmelen, F., Horrocks, I., & McGuinness, D. , L. Ontology languages for knowledge management. Springer, 2001
- Pinto, H. S., Martins, J. P., & Bente, A.. Ontology-based knowledge management: A literature review. *The Journal of Computer Information Systems*, 46(2), 23-32, 2009.