

UNIVERSIDADE DO MINHO
DEPARTAMENTO DE INFORMÁTICA

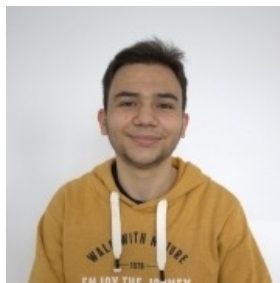


MESTRADO EM ENGENHARIA INFORMÁTICA

REPRESENTAÇÃO E PROCESSAMENTO DE CONHECIMENTO NA
WEB



Ana João Alves PG57505



Gonçalo Brandão PG57874

Maio de 2025

Conteúdo

1	Introdução	2
2	Ontologia	2
2.1	Arquitetura da Ontologia e Modelação Semântica	2
2.2	Metodologia de Povoamento e Processamento de Dados	2
2.3	Normalização de Dados	3
2.4	Estratégias de Qualidade e Consistência de Dados	3
2.5	Ontologia Final	3
3	Quiz	4
3.1	Desenvolvimento da Aplicação de Demonstração	4
3.2	Componentes Principais da Aplicação	4
3.3	Análise de Resultados e Métricas de Desempenho	4
4	Resultados	4
5	Conclusão	6
5.1	Considerações Técnicas e Escalabilidade	6
5.2	Fluxo de Processamento	6

1 Introdução

Este relatório apresenta uma análise detalhada do desenvolvimento de uma ontologia RDF/OWL abrangente sobre os Mundiais de Futebol, demonstrando a aplicação prática dos princípios da Web Semântica no domínio desportivo. O trabalho desenvolvido integra múltiplas fontes de dados estruturados para criar uma base de conhecimento rica e interconectada, capaz de representar semanticamente as complexas relações entre entidades do universo futebolístico mundial. A motivação para este projeto surge da necessidade crescente de estruturar e relacionar grandes volumes de dados desportivos de forma semanticamente rica, permitindo consultas sofisticadas e descoberta automática de conhecimento. Os Mundiais de Futebol, enquanto competição global de prestígio, oferecem um domínio ideal para demonstrar as capacidades das tecnologias da Web Semântica, dada a riqueza e diversidade dos dados envolvidos: desde informações sobre jogadores, equipas e estádios, até estatísticas detalhadas de partidas e torneios.

2 Ontologia

2.1 Arquitetura da Ontologia e Modelação Semântica

A ontologia desenvolvida fundamenta-se numa arquitetura robusta que captura a complexidade inerente ao domínio dos Mundiais de Futebol. A estrutura conceptual foi concebida para representar nove classes principais que constituem os pilares fundamentais da base de conhecimento: *Stadium* (instalações desportivas), *City* (localidades geográficas), *Country* (entidades nacionais), *Referee* (árbitros), *Coach* (treinadores), *Player* (jogadores), *Tournament* (competições), *Team* (equipas nacionais) e *Match* (partidas).

Esta modelação reflecte uma compreensão profunda das relações existentes no domínio futebolístico, onde cada entidade mantém conexões semânticas múltiplas e bidireccionais com outras entidades. Por exemplo, um jogador não existe isoladamente, mas mantém relações com equipas que representa, torneios em que participa, partidas em que joga e golos que marca. Esta interconnectividade é fundamental para permitir consultas sofisticadas e descoberta automática de padrões nos dados.

As propriedades de dados incluem identificadores únicos, nomes normalizados, informações temporais como datas de nascimento e realização de eventos, capacidades de estádios, resultados quantitativos e ligações externas para recursos como páginas da Wikipedia. Paralelamente, as propriedades de objeto estabelecem as relações semânticas cruciais entre entidades, incluindo localização geográfica, participação em competições, representação nacional e desempenho desportivo.

Classe	Descrição
Stadium	Instalações desportivas
City	Localidades geográficas
Country	Entidades nacionais
Referee	Árbitros
Coach	Treinadores
Player	Jogadores
Tournament	Competições
Team	Equipas nacionais
Match	Partidas

Tabela 1: Classes principais da ontologia

2.2 Metodologia de Povoamento e Processamento de Dados

O processo de povoamento da ontologia foi estruturado em duas fases distintas mas complementares, cada uma implementada através de scripts especializados que garantem a integridade e consistência da base de conhecimento resultante.

A primeira fase, implementada através do script `worldcup_povoar.py`, concentra-se no povoamento inicial da ontologia mediante o processamento sequencial de múltiplos ficheiros CSV contendo dados estruturados sobre diferentes aspetos do domínio futebolístico. Este script processa sistematicamente informações sobre estádios, incluindo as suas características físicas como capacidade e localização geográfica, criando simultaneamente instâncias para cidades e

países quando necessário. O processamento de dados sobre árbitros incorpora a distinção entre géneros, enquanto a gestão de informações sobre treinadores inclui tratamento especial para valores ausentes ou não aplicáveis.

O processamento de dados sobre jogadores revela particular complexidade, uma vez que envolve a determinação automática de posições baseada em campos booleanos e a gestão de listas de torneios em que cada jogador participou. Os torneios são processados com atenção especial às suas características temporais e geográficas, incluindo a identificação automática entre competições masculinas e femininas. As equipas requerem tratamento especial devido à necessidade de criar instâncias separadas para representações masculinas e femininas, cada uma com ligações específicas para recursos externos.

A segunda fase, operacionalizada pelo script `worldcup_relacoes.py`, estabelece as relações complexas entre entidades previamente criadas. Este processo utiliza a biblioteca `rdflib` para manipulação direta do grafo RDF, permitindo um controlo fino sobre a criação de triplos semânticos. O processamento de escalões estabelece vínculos bidirecionais entre jogadores e equipas, complementados por ligações entre equipas e países que representam, bem como a sua participação em torneios específicos.

2.3 Normalização de Dados

A função de normalização implementada constitui um elemento crucial para garantir a consistência dos identificadores RDF/OWL:

Listing 1: Função de normalização de strings

```
def norm(x):  
    return x.replace("-", "_").replace(" ", "").replace("'", "") \  
           .replace("(", "").replace(")", "").replace("—", "")  
    # ... outras substituições
```

Esta implementação assegura que todos os identificadores são compatíveis com os requisitos sintáticos do RDF/OWL, eliminando caracteres que poderiam causar problemas de parsing ou interpretação.

2.4 Estratégias de Qualidade e Consistência de Dados

A garantia de qualidade dos dados constitui um aspeto fundamental do desenvolvimento desta ontologia, implementada através de múltiplas estratégias complementares que asseguram a consistência, completude e utilizabilidade da base de conhecimento resultante.

A normalização de dados representa uma preocupação central, operacionalizada através de uma função especializada que implementa transformações extensivas de strings, substituindo caracteres especiais e espaços por underscores para garantir compatibilidade total com identificadores RDF/OWL. Esta abordagem sistemática previne inconsistências que poderiam comprometer a integridade referencial da ontologia.

A gestão de identidades únicas constitui outro pilar fundamental da estratégia de qualidade, implementada através de algoritmos que combinam informações contextuais como género e torneio para gerar identificadores únicos e persistentes. Esta estratégia é particularmente importante quando se considera a necessidade de distinguir entre entidades aparentemente similares mas contextualmente distintas, como equipas masculinas e femininas do mesmo país.

O tratamento robusto de dados inconsistentes ou ausentes é operacionalizado através de estratégias de fallback que garantem a continuidade do processamento mesmo na presença de informações incompletas. Campos marcados como “not applicable” ou “not available” são tratados de forma consistente, evitando a introdução de dados inválidos na ontologia.

2.5 Ontologia Final

- Número de classes: 9
- Número de object properties: 23
- Número de data properties: 22
- Número de indivíduos: 14227
- Número de triplos: 161991

3 Quiz

3.1 Desenvolvimento da Aplicação de Demonstração

A aplicação web desenvolvida em Flask constitui uma demonstração prática e interativa das capacidades da ontologia, materializando o valor da base de conhecimento através de uma experiência de utilizador envolvente e educativa. Esta aplicação não representa meramente um exercício técnico, mas sim uma validação real da utilidade e acessibilidade do conhecimento estruturado.

A arquitetura da aplicação fundamenta-se numa integração direta com o GraphDB através de consultas SPARQL, demonstrando como ontologias podem ser operacionalizadas em aplicações práticas. O sistema implementa múltiplos modos de dificuldade, desde níveis introdutórios até desafios especializados, incluindo um modo de “morte súbita” que testa o conhecimento sob pressão temporal.

A geração dinâmica de perguntas baseia-se em consultas SPARQL que extraem informações da ontologia em tempo real, garantindo que o conteúdo do quiz reflecte sempre o estado atual da base de conhecimento. Os tipos de pergunta implementados incluem escolha múltipla sobre características de torneios, associações entre jogadores e posições, questões verdadeiro/falso sobre estatísticas de golos, e desafios mais complexos que requerem conhecimento profundo das relações entre entidades.

3.2 Componentes Principais da Aplicação

A aplicação estrutura-se em torno de várias componentes fundamentais que trabalham em conjunto para proporcionar uma experiência interativa rica:

- **Configuração Inicial:** Estabelece a ligação com o endpoint GraphDB e define os diferentes modos de jogo disponíveis
- **Funções de Extração:** Implementam consultas SPARQL especializadas para obter diferentes tipos de informação da ontologia
- **Rotas Flask:** Gerem o fluxo da aplicação desde a página inicial até à apresentação de resultados
- **Gestão de Sessão:** Mantém o estado do utilizador ao longo da interação com o sistema

3.3 Análise de Resultados e Métricas de Desempenho


A avaliação quantitativa da ontologia desenvolvida revela resultados que confirmam tanto a abrangência quanto a eficiência dos processos implementados. A contagem total de triplos RDF gerados fornece uma métrica fundamental sobre a densidade da base de conhecimento, enquanto a análise de propriedades não utilizadas confirma a eficiência do design ontológico.

A distribuição de entidades por classe revela um equilíbrio apropriado entre diferentes tipos de informação, indicando uma cobertura abrangente do domínio sem redundância excessiva. A densidade de relações estabelecidas demonstra a riqueza semântica da ontologia, confirmando que as entidades não existem isoladamente mas integram uma rede coesa de conhecimento interconectado.

As métricas de desempenho da aplicação web, incluindo tempos de resposta para consultas SPARQL e taxas de sucesso na geração de perguntas, confirmam a viabilidade prática da solução desenvolvida. A variabilidade e qualidade das perguntas geradas automaticamente demonstram a riqueza da base de conhecimento e a eficácia das consultas implementadas.

4 Resultados

Nesta secção serão demonstradas algumas páginas do quiz mencionado anteriormente.



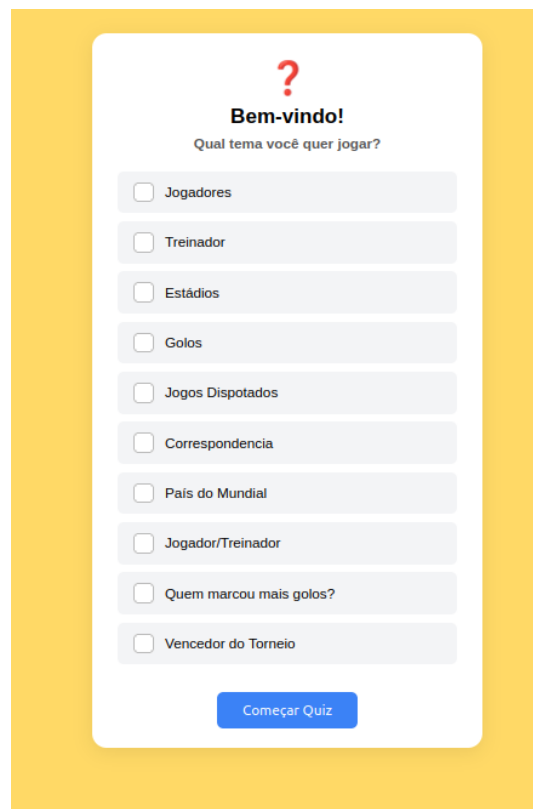
A screenshot of a quiz selection screen with a yellow background. At the top, there is a trophy icon and the text "BEM VINDO AO QUIZ DO MUNDIAL". Below this, it says "Escolha o modo de jogo:". There are four radio button options: "Fácil" (with a trophy icon), "Mediano", "Expert", and "Morte Súbita". At the bottom, there is a blue button labeled "Escolher Temas".

BEM VINDO AO QUIZ DO MUNDIAL
Escolha o modo de jogo:

- ☐ Fácil 🏆
- ☐ Mediano
- ☐ Expert
- ☐ Morte Súbita

[Escolher Temas](#)

Figura 1: Níveis do quiz



A screenshot of a quiz topic selection screen with a yellow background. At the top, there is a red question mark icon and the text "Bem-vindo!". Below this, it says "Qual tema você quer jogar?". There are ten radio button options: "Jogadores", "Treinador", "Estádios", "Golos", "Jogos Disputados", "Correspondencia", "País do Mundial", "Jogador/Treinador", "Quem marcou mais golos?", and "Vencedor do Torneio". At the bottom, there is a blue button labeled "Começar Quiz".

Bem-vindo!
Qual tema você quer jogar?

- ☐ Jogadores
- ☐ Treinador
- ☐ Estádios
- ☐ Golos
- ☐ Jogos Disputados
- ☐ Correspondencia
- ☐ País do Mundial
- ☐ Jogador/Treinador
- ☐ Quem marcou mais golos?
- ☐ Vencedor do Torneio

[Começar Quiz](#)

Figura 2: Temas do quiz

Que treinador treinou Colômbia no Mundial de 1998?

A György Orth B Javier Clemente

C Hernán Darío Gómez D Enzo Bearzot

Confirmar resposta

Pontuação atual: 0

Figura 3: Pergunta de tema de treinadores

Associe cada atleta à sua posição

Mounir El Hadjoui	goalkeeper
Pierre Womé	goalkeeper
Willy Huber	goalkeeper
Andressinha	goalkeeper

Enviar Respostas

Pontuação atual: 0

Figura 4: Pergunta de Correspondência

5 Conclusão

5.1 Considerações Técnicas e Escalabilidade

O desenvolvimento desta ontologia incorporou desde o início preocupações com escalabilidade e manutenibilidade que são fundamentais para sistemas de conhecimento destinados a evoluir e crescer ao longo do tempo. A utilização de estruturas de dados eficientes e processamento sequencial permite suportar volumes significativos de dados sem degradação substancial do desempenho.

A modularidade da arquitetura, expressa através da separação entre scripts de povoamento e estabelecimento de relações, facilita a manutenção e extensão do sistema. Esta abordagem permite atualizações incrementais da base de conhecimento sem necessidade de reproprocessamento completo, um aspeto crucial para sistemas que devem incorporar novos dados regularmente.

As estratégias implementadas para gestão de memória e otimização de consultas SPARQL garantem que a aplicação mantém desempenho responsivo mesmo com o crescimento da base de conhecimento. A utilização de índices apropriados e estruturação eficiente dos dados suporta consultas complexas sem impacto significativo na experiência do utilizador.

5.2 Fluxo de Processamento

O processamento dos dados segue uma sequência bem definida que garante a integridade e consistência da ontologia resultante:

1. **Inicialização:** Cópia e preparação do ficheiro base da ontologia
2. **Povoamento:** Processamento sequencial dos ficheiros CSV com dados das entidades
3. **Normalização:** Aplicação sistemática de regras de limpeza e padronização
4. **Instanciação:** Criação das entidades RDF/OWL na ontologia

5. **Relacionamento:** Estabelecimento das relações semânticas entre entidades
6. **Validação:** Verificação da consistência e completude da ontologia
7. **Serialização:** Exportação final para formato Turtle

O sucesso deste projeto reside não apenas na implementação técnica competente, mas na demonstração clara de como as tecnologias da Web Semântica podem transformar dados brutos em conhecimento estruturado e consultável. A ontologia dos Mundiais de Futebol desenvolvida representa mais do que uma coleção de dados; constitui um exemplo paradigmático de como o conhecimento humano pode ser formalizado e disponibilizado de forma que tanto humanos quanto máquinas possam compreender e utilizar eficazmente. Este trabalho confirma que a Web Semântica, quando aplicada com rigor metodológico e visão estratégica, oferece ferramentas poderosas para a gestão do conhecimento no século XXI, contribuindo para um futuro onde a informação é não apenas acessível, mas verdadeiramente compreensível e utilizável por sistemas inteligentes.