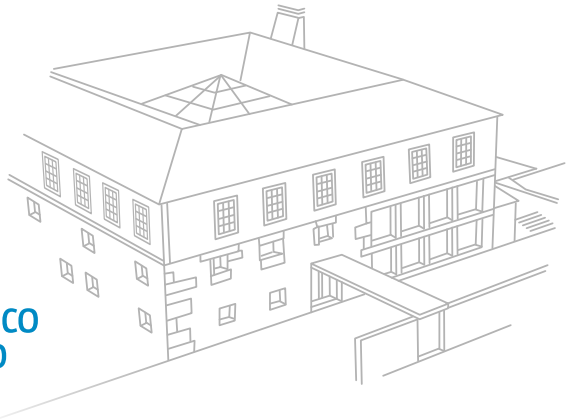


ESTGF | POLITÉCNICO DO PORTO

ESCOLA SUPERIOR DE TECNOLOGIA E GESTÃO

Uma nova abordagem para partilha de informação no contexto de integração de plataformas colaborativas: recurso a Linked Data

DESIGNAÇÃO DO MESTRADO

Mestrado Engenharia Informática

AUTOR

Fábio Joel Neves Alves

ORIENTADOR(ES)

Professora Doutora Carla Pereira

Mestre Cristóvão Sousa

ANO

2014

Agradecimentos

No final de mais uma etapa do meu percurso académico e de mais um objetivo da minha vida atingido, não posso deixar de mencionar um conjunto de pessoas e instituições que para mim foram importantes.

À Professora Carla Pereira e Professor Cristóvão Sousa, orientadora e coorientador desta dissertação, por toda a atenção e compreensão que tiveram para comigo em todas as fases da elaboração desta dissertação. Os altos e baixos que surgiram no decorrer desta dissertação, levou-me a perceber que além de dois orientadores da dissertação, tinha aqui também dois amigos que nunca mais poderei esquecer.

Não posso também deixar de agradecer a todos os membros da Unidade de Engenharia e Sistemas de Produção (UESP) do INESC Tec, todo o apoio no meu percurso académico que já perdura desde a minha licenciatura.

À minha esposa Marlene, que teve um papel fundamental na minha motivação e compreensão nas horas investidas no desenvolvimento da dissertação.

À minha família.

Muito Obrigado!!!!!!

Resumo PT

Esta dissertação, integrada no contexto da gestão de conhecimento e informação em redes colaborativas, teve como principal objetivo a construção de uma solução que permitisse a partilha de informação entre plataformas de dados estruturados. Surge na sequência de um projeto europeu, já concluído, designado por H-Know, dando seguimento a necessidades futuras identificas e ainda não colmatadas. Assim, em termos genéricos, esta dissertação foca-se na necessidade de melhorar a colaboração e partilha de informação na Internet, sendo estudados métodos, técnicas e ferramentas que permitam responder a tal necessidade.

Como resultado do projeto H-Know surgiu ainda uma plataforma *Web*, colaborativa, com funcionalidades avançadas de gestão de conhecimento e informação na área de reabilitação de edifícios. Esta plataforma H-Know, sendo um sucesso de implementação, surge aqui como principal inspiração para este trabalho de investigação, não só orientado à plataforma H-Know, mas também à criação de serviços genéricos que permitam o enriquecimento de informação em plataformas *Web* colaborativas de dados estruturados. O processamento de ontologias e dados estruturados, disponibilizados quer através de ficheiros RDF e/ou *triple stores* através de SPARQL *endpoints,* permitiu ao serviço de pesquisas desenvolvido melhorar o resultado das mesmas em plataformas cliente.

Palavras-chave: Gestão de Informação, Gestão de Conhecimento, Colaboração, Redes Colaborativos*, Knowledge Discovery*, *Linked Data*.

Resumo EN

This dissertation, within the context of knowledge and information management in collaborative networks, had as main goal the construction of a solution that would allow the information sharing between platforms of structured data. It arises as a result of a European Project referred to as H-Know, which is now already closed, following future needs which were identified but still not yet solved. In general, this dissertation focuses on the need to improve collaboration and information sharing on the Internet, by studying methods, techniques and tools which allow responding to such need.

As a result of the H-Know project, a collaborative *Web* platform also emerged, with advanced features for managing knowledge and information in the area of the rehabilitation of buildings. This H-Know platform, which was an implementation success, arises here as the main inspiration for this research work, not only oriented to this platform, but also to the creation of generic services that allow the enrichment of information in collaborative *Web* platforms of structured data. Ontologies processing and structured data, available either through RDF files and/or triple stores via SPARQL *endpoints*, allowed the research services designed to improve their outcomes in client platforms.

Keywords: Information Management, Knowledge Management, Collaboration, Collaborative networks*, Linked Data*.´

Índice

Índice Figuras

Índice Tabelas

Lista de Abreviaturas

URL - Uniform Resource Locator

URI - Uniform Resource Identifier

XML - eXtensible Markup Language

FMSLR - Fundación Santa Maria La Real

CAPEB - Confédération de l'Artisanat et des Petites Entreprises du Bâtiment

RDF - *Resource Description Framework*

OWL - Web Ontology Language

W3C - World Wide Web Consortium

API - Application programming interface

SOA - Service-oriented architecture

ERP - Enterprise resource planning

SCM - Supply chain management

CRM - Customer relationship management

VE – *Virtual Enterprise*

VO – *Virtual Organization*

VBE – *Virtual Organization Breeding Environment*

XML - *Extensible Markup Language*

W3C – *World Wide Web Consortium*

FOAF – *Friend of a Friend*

SIOC – *Semantically-Interlinked Online Communities*

DCMI – *Dublin Core*

VOID - Vocabulary of interLinked Datasets

SKOS - *Simple Knowledge Organization System*

CML – *Corporate Language Management*

CMS – *Content Management System*

PPT - PoolParty Thesaurus Server

PPX - PoolParty Extractor

PPS - PoolParty Semantic Search

PPP - PoolParty PowerTagging

PPI - PoolParty Semantic Integrator

Capítulo 1

Introdução

1.1 Contextualização

Atualmente a falta de informação não é um problema que se coloque. A problemática passa pelo seu armazenamento e recuperação de forma simples e clara, satisfazendo as necessidades dos utilizadores. Dada a relevância deste assunto no contexto socioeconómico atual é realizada anualmente uma conferência, SWCS (*Semantic Web Collaborative Spaces)*, onde são apresentadas e discutidas novas abordagens para ajudarem na partilha de informação em redes colaborativas. Esta problemática e o projeto H-Know surge como ponto de partida para esta dissertação.

Segundo o consórcio do projeto H-Know, este projeto teve como objetivos a criação de uma metodologia de suporte à partilha de conhecimento entre parceiros da construção civil em especial para pequenas e médias empresas que têm o seu modelo de negócio direcionado para a reabilitação de edifícios antigos nos quais a sua estrutura e herança cultural são de grande importância. O segundo objetivo passou pela criação de uma plataforma *Web* que para além de suportar a metodologia implementada, ajuda-se no melhoramento da cooperação entre PME´s[1] [1].

Os conceitos de redes colaborativas, gestão de conhecimento, recuperação de informação e *Linked Data*, aqui introduzidos, são conceitos que se encontram na base do projeto H-Know, que potenciou o desenvolvimento desta tese. Assim, torna-se evidente a utilização do projeto, como referencia nesta dissertação, não só nos capítulos de estado da arte, como também nos cenários de teste definidos.

[1] Pequenas e médias empresas

1.2 Conceitos base

Segundo [2], uma rede (ou grafo) é um conjunto de vértices conectados através de relações que podem representar pontos de ligação entre pessoas, comunicações, representações físicas entre outros. Entretanto, o conceito tem vindo a ser diversificado, levando a que neste trabalho seja adotada a definição proposta por [3]. Segundo [3] as redes colaborativas são constituídas por um conjunto de entidades (pessoas e/ou instituições), autónomas, heterogéneas, geograficamente separadas que possuem objetivos, culturas e recursos diferentes, mas que possuem interações suportadas por computadores para a resolução de pontos de interesse comuns. Em termos genéricos, o presente trabalho pretende dar suporte à operacionalização de redes colaborativas com recurso ao conceito dos espaços colaborativos. Espaços colaborativos são por definição [4] virtuais, integrados em plataformas (*Web-based*) de dados estruturados.

As redes colaborativas pautam-se por atividades distribuídas que requerem elevada interação social, em que os processos de gestão de informação e partilha de conhecimento assumem particular relevo. Em [5], os autores afirmam que foram as empresas do ramo industrial que levantaram o interesse pela gestão de conhecimento através da recolha da sua própria informação, partilhando-a com clientes e acionistas. Esta necessidade surge pelo facto das organizações ganharem consciência, por um lado, que todos os intervenientes da organização são importantes e necessitam de se sentir como parte desta, tornando assim a criatividade e o interesse pela organização cada vez maior. Por outro lado, a partilha de informação possibilita o relacionamento com outra informação, levando assim ao desenvolvimento de novas técnicas e tecnologias, dando origem a mais e melhor inovação [5].

Atualmente com os novos desafios para a colaboração, onde a aprendizagem colaborativa, inovação e criatividade são valorizados pela sociedade, são um desafio, que obriga à existência de mecanismos que permitam a criação de relações entre dados pessoas, tal como a abordagem *Linked Data.*

O conceito *Linked Data* surge em [6], que o apresenta como um *standard* para a publicação de dados na Internet. Com este conceito a informação torna-se referenciável permitindo assim uma mais fácil reutilização de informação, passando o esforço de desenvolvimento para a construção de aplicações que “consomem” dados. Estes dados podem ser procurados em diversas fontes de dados dispersas geograficamente, como documentos e páginas *Web* que são processadas por serviços que os disponibilizam de

acordo com o *standard*. Assim torna-se possível obter grafos de representação de conhecimento em que as arestas são os URI que identificam a informação.

1.3 Objetivos e motivação

Esta dissertação, tal como já foi referido, surge na sequência de um projeto europeu, já concluído, designado por H-Know[2]. A problemática estudada no âmbito deste projeto é bastante atual, e resulta da necessidade de melhorar a colaboração e partilha de informação de um determinado domínio através da Internet. Assim, nesta dissertação são estudados métodos, técnicas e ferramentas que permitam contribuir para satisfazer seguintes as necessidades: a) falta de informação do domínio da reabilitação tendo em conta a herança cultural, área geográfica e língua; b) falta de conhecimento da existência de fornecedores de materiais e/ou mãos de obra especializada para determinadas operações; c) combater a dispersão de informação como legislação, questões ambientes, entre outros; d) potenciar as parcerias entre empresas e organizações do mesmo ramo; identificadas durante o decorrer do projeto. Como resultado do projeto H-Know surgiu uma plataforma para a *Web*, colaborativa, com funcionalidades avançadas de gestão de conhecimento na área de reabilitação de edifícios. A plataforma H-Know foi um sucesso de implementação, a utilização foi garantida pelos parceiros responsáveis pela elaboração do modelo de negócio e respetiva implementação. O *feedback* de todos os parceiros, industriais e investigação, foi muito importante e vinculativo. Assim, surgiu esta oportunidade de investigação não só orientada à plataforma H-Know, mas também à criação de serviços genéricos que permitam o enriquecimento de informação em plataformas *Web* colaborativas de dados estruturados, considerando um domínio específico. A participação neste projeto, permitiu-nos que o mesmo viesse a ser usado nesta dissertação como uma excelente estudo de caso e cenário de implementação do serviço desenvolvido.

Assim, de forma sintetizada os objetivos desta dissertação são:

1. Definição de uma solução de engenharia que facilite a partilha de informação estruturada entre plataformas de dados estruturados;
2. Seleção de um conjunto de ferramentas e tecnologias *Linked Data OpenSource*, a serem utilizadas em 3;

[2] http://h-know.eu/

3. Construção, teste e validação de um serviço de pesquisa de informação estruturada entre plataformas colaborativas de dados estruturados.

Os principais resultados são:

1. Arquitetura de uma solução *Linked Data*, que permite a partilha de informação entre plataformas de dados estruturados;
2. Desenvolvimento de um motor de pesquisa, na forma de serviço, que permita o acesso simplificado e intuitivo à informação efetivamente relevante para os utilizadores finais.

1.4 Estrutura da tese

Este documento encontra-se estruturado em seis capítulos dos quais o primeiro é uma apresentação e contextualização do trabalho.

O segundo capítulo apresenta uma breve análise do estado da arte das redes colaborativas, citando o conceito, manifestações e variantes.

No terceiro capítulo é efetuada uma contextualização da necessidade de integração de informação entre plataformas empresariais. É introduzido o conceito *Linked Data,* e efetuado um levantamento de abordagens e plataformas que cumprem os princípios do *Linked Data.*

O capítulo quatro apresenta a solução de engenharia desenvolvida, principal resultado desta dissertação.

O capítulo cinco, carateriza-se pela descrição do procedimento de teste e validação do *semantic search service proposto no capítulo anterior.* São também apresentados e analisados os resultados ***"as-is"*** e ***"to-be"*** obtidos na execução dos cenários de teste definidos.

Capítulo 2

Redes Colaborativas

2.1 As Redes colaborativas no contexto de partilha de Informação

No ambiente competitivo atual em que nos encontramos quer as organizações industriais quer de serviços têm ao longo dos anos sentido a necessidade de se reestruturarem. Segundo [9], as organizações para saírem com sucesso desta fase, até 2020 devem efetuar alterações significativas adaptando-se a novos modelos de negócio, estratégias, questões governamentais e tecnológicas. Com isto, no Conselho de investigação realizado em Washington em 1998 [8], foram identificados seis grandes desafios propostos para as organizações tendo em conta o período até 2020. Os desafios são:

1. Alcançar a simultaneidade de operações;
2. Integração humana e recursos técnicos de modo a aumentar a força de trabalho com a satisfação;
3. Transformar rapidamente a informação recolhida de diversas fontes em conhecimento importante na tomada de decisões;
4. Rápida reorganização das empresas de modo a responder rapidamente a necessidades e oportunidades;
5. Reduzir os desperdícios bem como tornar o impacto ambiental perto de nulo;
6. Desenvolver processos inovadores de produção orientados ao produto eliminando a visão a grande escala.

Em teoria estes desafios indicam que as novas estruturas organizacionais, modelos de negócio e processos devem permitir às organizações uma maior dinâmica na mudança e nas operações. É neste contexto que segundo [9], a colaboração tem um papel essencial, nomeadamente entre as pequenas e médias empresas que possuem recursos e

competências limitadas, mas que através de uma colaboração organizada e orientada podem dinamicamente restruturar-se diversificando as suas competências e áreas de intervenção. Neste mesmo artigo, são identificados quatro pontos-chave responsáveis pelo sucesso num processo de colaboração [9]:

Rede: Envolve basicamente a comunicação e informação partilhada em benefício mútuo. Um caso prático e típico de uma rede, é quando um grupo de entidades partilham informação sobre a sua experiência com a utilização de uma determinada ferramenta. Esta informação pode ser importante, contudo não é divulgada informação crucial "chave do negócio", como para quê que esta é utilizada. Através destes "gestos" de benefício mútuo, torna-se possível estabelecer relações de confiança e introduzir nas pessoas o conceito de redes.

Coordenação da rede: Esta atividade encontra-se diretamente relacionada com a anterior, na medida em que vai garantir uma maior eficiência nos resultados obtidos. A coordenação deve possuir um papel harmonizador e concertado com a rede. Os responsáveis por esta atividade, devem orientar a divulgação de informação, tempos e maximizar o impacto da informação divulgada.

Cooperação: Muitos autores, em diversas áreas de investigação, fazem referência à cooperação e colaboração como sendo um só [10] [11]. Contudo, neste contexto a cooperação é bem distinta da colaboração, na medida em que passa pela partilha e troca de recursos para alcançarem um objetivo semelhante.

Colaboração: Ao contrário da cooperação, a colaboração parte do princípio que existe partilha de informação, recursos e responsabilidades como um todo. Enquanto na cooperação se pressupõe a partilha de tarefas e ou recursos. Um caso prático de colaboração acontece, por exemplo, quando um grupo de investigadores se junta em torno de um determinado problema e em conjunto criam sinergias para a sua resolução.

Assim, podemos evidenciar que uma rede colaborativa possui um ciclo de vida constituído possivelmente pelos cincos estados apresentados na Figura 1 [12]

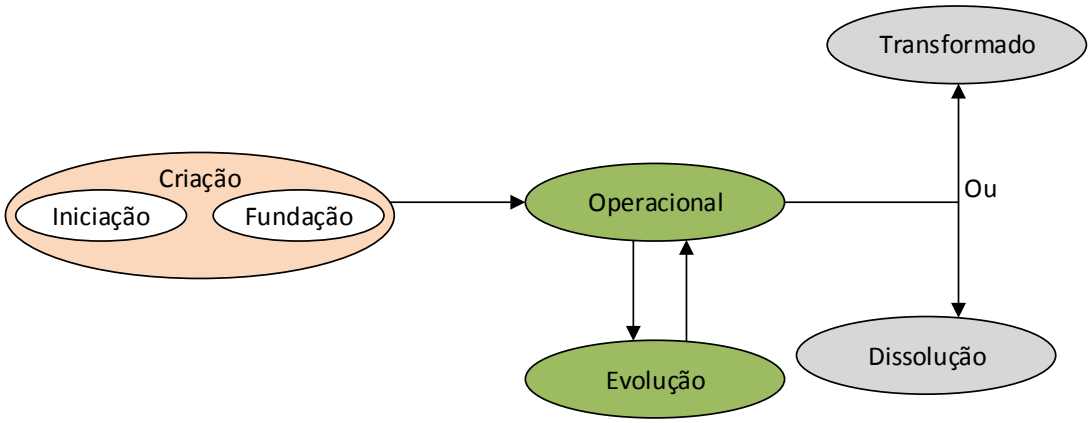


Figura 1: Ciclo de vida de uma rede colaborativa *[6]*

Criação: Este estado existe quando se dá início ao processo de criação de uma rede colaborativa. Este processo pode ser dividido em duas fases:

(i) a fase de iniciação e recrutamento, altura em que se faz um delineamento da estratégia e planeamento;

(ii) (ii) quando existe realmente a iniciação e arranque.

Operação: Fase em que a rede está operacional, ou seja, encontra-se com um número de intervenientes estável que garante o funcionamento da mesma.

Evolução: Este estado acontece quando existe alguma reformulação de objetivos ou do enquadramento desta rede. Pode levar a pequenas alterações de membros, papeis ou alteração de princípios.

Dissolução: Quando a rede deixa de fazer sentido.

Transformação: Quando os princípios foram importantes para que os objetivos fossem atingidos. Assim, pode aproveitar-se o que de melhor esta rede ofereceu para a criação de uma nova rede com objetivos a definir.

2.2 Manifestações ou variantes das redes colaborativas

Existem inúmeras manifestações de redes colaborativas que incluem: organizações virtuais, empresas virtuais, cadeias de valor dinâmicas e comunidades virtuais de profissionais. Cada uma destas categorias encontra-se definida em [13].

***Virtual Enterprise* (VE)**: Aliança temporária de empresas para partilharem competências e recursos com o objetivo de responderem melhor a oportunidades de negócio. Aliança suportada por redes informáticas.

***Virtual Organization* (VO)**: Conceito similar a VE, contudo a aliança não é utilizada única e simplesmente na obtenção de lucros. Assim, as VE, podem ser conhecidas como casos particulares das VO.

Dynamic Virtual Organization: Tipicamente são conhecidas como VO, contudo a sua existência é curta, uma vez que são constituídas e "destruídas" num curto espaço de tempo. Estas alianças existem segundo um propósito, normalmente com um âmbito muito pequeno.

Extended Enterprise: É um conceito tipicamente aplicado a empresas que definem que as suas fronteiras de partilha de informação terminam nos seus fornecedores.

***VO Breeding Environment* (VBE)**: Neste caso em particular, é uma organização ou conjunto de organizações conhecidas por possuírem uma rede de contactos que podem ser usadas por empresas de modo a estabelecerem cooperações, colaborações ou interoperabilidade de infraestruturas. Um exemplo são as incubadoras de empresas, *start-ups.*

No contexto deste trabalho, as três primeiras categorias apresentadas, são aquelas que nos vão dar apoio no decorrer desta dissertação. O apoio às organizações para o fortalecimento de parcerias potenciando o desenvolvimento de projetos conjuntos, apresentação de portefólios, apresentação de novos produtos, distribuição de competências, são algumas das áreas onde as categorias referenciadas se podem enquadrar. [14] afirma ainda que os espaços colaborativos podem ser utilizados como mediadores na partilha de informação em redes colaborativas.

Assim, os espaços colaborativos foram analisados, de modo a materializar uma rede colaborativa no suporte às suas operações básicas. Segundo [13], os espaços colaborativos são estruturas de suporte às redes colaborativas compostas por um conjunto de funcionalidades que visam o agrupamento de um conjunto de pessoas/organizações que trabalham para alcançar o mesmo objetivo. De acordo com [14], os espaços colaborativos, tal como as próprias redes colaborativas, possuem um ciclo de vida constituído por quatro estados que se encontram apresentados na Figura 2.

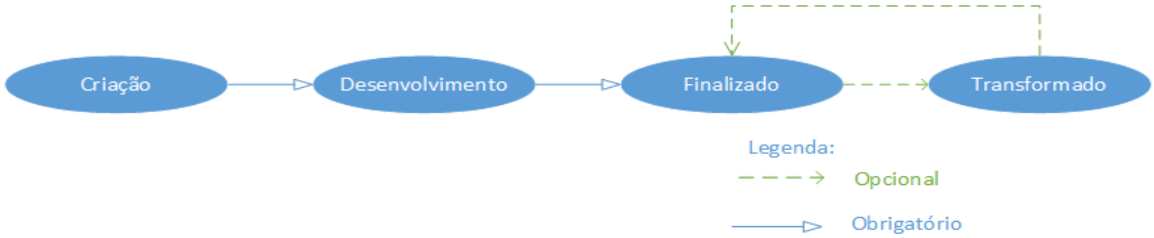


Figura 2:Ciclo de vida de um espaço colaborativo

1) O estado “criação”, tal como o próprio nome indica, é quando o espaço é criado e este não é alterado até à definição dos objetivos e formação da equipa de trabalho;
2) Uma vez os objetivos e a equipa inicial definida, o estado passa para “desenvolvimento”, neste estado é expectável que se desenvolvam todos os trabalhos, relatórios e informação necessária ao cumprimento dos objetivos;
3) Obtendo os resultados esperados, o espaço é “finalizado”, com isto as funcionalidades de edição, adição e utilização são bloqueadas, passando a ser possível, única e simplesmente, a consulta de informação registada neste;
4) O último estado “transformação”, é aquele que motiva a afirmação apresentada anteriormente - reutilização de informação, isto porque é através deste estado que se torna possível pegar em informação relevante e não confidencial reaproveitando-a para um novo espaço. De acordo com [14], é com base neste estado que os resultados de um espaço colaborativo podem servir para a definição de objetivos de outro.

Nesta fase o leitor pergunta-se: “mas então o espaço colaborativo é construído segundo um ciclo? É possível usar os espaços colaborativos e não respeitar os ciclos?”, a resposta é “Sim”. É certo que os espaços colaborativos podem ser criados, desenvolvidos e finalizados. Contudo, pretende-se neste trabalho dar especial ênfase a reutilização de informação que pode essencialmente ser obtida aquando da passagem ao estado “transformação”.

Capítulo 3

Recurso a *Linked Data* na integração de Plataformas Colaborativas

3.1 Contextualização

"*The Semantic Web is an extension of the current Web in which information is given well-defined meaning, better enabling computers and people to work in cooperation.*"

Tim Berners-Lee, James Hendler, Ora Lassila (2001)

A Internet semântica, é uma extensão da Internet atual, e não pode ser encarada como se fosse desenvolvida a par da atual, mas sim como uma extensão. Baseia-se no conteúdo atual e nas estruturas atuais da Internet, mas adiciona-lhe novos recursos que permitem tornar os dados processáveis por máquinas, recorrendo a padrões nas relações que torna as associações mais ricas [16]. Esta descrição adicional, permite por exemplo que as páginas *Web* forneçam informação mais detalhada a motores de busca, de modo a que as pesquisas sejam melhor direcionadas. A possibilidade de reutilização de informação e o processamento de informação através de máquinas, introduz na comunidade científica um conceito designado por interoperabilidade semântica[17]. Segundo uma diretiva da união europeia para *software*[3], o conceito de interoperabilidade é definido da seguinte forma: "*the ability to exchange information and use the information which has been exchanged*".

Assim, adaptando a definição ao contexto deste trabalho, podemos dizer que a interoperabilidade semântica é a capacidade de trabalhar com dados distribuídos por diversas fontes, considerando a heterogeneidade semântica e sintática entre as diferentes fontes.

Em [18], a interoperabilidade é apresentada como um requisito base para os sistemas de informação modernos. No livro, o autor procurou responder a algumas questões que considerou importantes e ainda bastante atuais, como por exemplo: Quais são os requisitos fundamentais para a interoperabilidade ao longo do tempo? Como podemos perceber o

[3] European Union Software Directive, see Council Directive 91/250/EEC

âmbito da interoperabilidade? O que tem sido feito, para a interoperabilidade em sistemas de informações? O autor identifica também dois objetivos importantes que justificam o estudo da interoperabilidade semântica e que se encontram totalmente alinhados com a atualidade, que são:

- Desenvolvimento tecnológico de intercomunicação através da Internet.
- Diversidade de fontes de informação distribuídas pela Internet, bem como a sua variedade.
- Aumento da especialização no trabalho, que obriga a um aumento na reutilização de dados, permitindo a criação de conhecimento partilhável e reutilizável.

[*32*] defende que no contexto atual algumas das questões colocadas por Sheth em 1998, já se encontram ultrapassadas, nomeadamente as de carácter mais tecnológico. Desde então, os sistemas de informação passaram a ser orientados a serviços (SOA), tendo sido já desenvolvidos alguns *standards* de interoperabilidade de dados, e já se encontram implementados e aceites pela comunidade, como é o caso do XML (eXtensible Markup Language).

No seguimento destes desenvolvimentos, recentemente a W3C (*World Wide Web Consortium*)[4], aprovou duas estruturas lógicas(baseadas em XML), de dados que possibilitam a criação de informação mais rica, autónoma e mais acessível que são o OWL (*Web* Ontology Language)[5] e RDF (Resource Description *Framework*)[6]: Estas estruturas veem ajudar as linguagens definidas pela W3C, como *markup languages* através da extensão à vertente de representação de conhecimento. De modo genérico o RDF é uma *framework* que permite estabelecer relações entre dados, enquanto que o OWL além das capacidades do RDF permite especificar restrições entre dados e relações. Segundo [16], o "sonho" da W3C, passa por potenciar o desenvolvimento de abordagens e/ou tecnologias que permitam que a informação seja de igual modo percetível e processável quer por pessoas como por máquinas.

[4] http://www.w3.org/
[5] http://www.w3.org/TR/owl-features/
[6] http://www.w3.org/RDF/

3.1.1 RDF

O RDF, segundo [20], é considerado o *standard* mais relevante na representação e interoperabilidade de dados estruturados. Neste mesmo artigo, os autores referenciam [21], para apresentarem as recomendações da *W3C que* identificam o RDF como o *standard* mais completo em termos de sintaxe. Esta caraterística é vantajosa no sentido em que torna o modelo de dados claro e de fácil interpretação. A terminologia RDF encontra-se associada aos *statements (subject, predicate, object)*, conhecidos como triplos RDF de representação de dados.

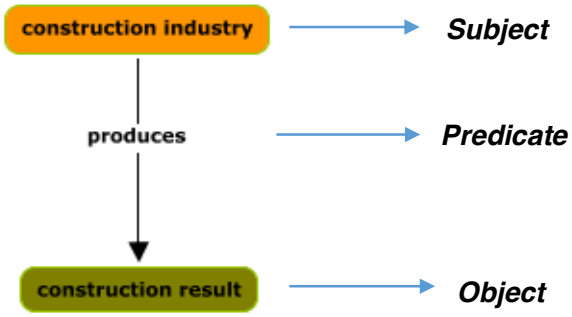


Figura 3:Triplo RDF

Atualmente existem três notações distintas de representação em RDF: RDF *Triples*, RDF *graphs* e RDF/XML que podem ser utilizadas de acordo com as plataformas e tipos de sistemas em que se encontram inseridas. Na tabela 1, podemos encontrar a representação em RDF *triple* e RDF/XML, a representação RDF *graphs* pode ser visualizada na Figura 3.

Representação	Tipo Representação
Contruction Industry a Col:produces rdf:about construction result	Notation 3
<rdf:Description rdf:about=construction_industry> <col:produces> construction_result <col:produces> </rdf:Description>	RDF/XML

Tabela 1:Exemplos de tipos de representação RDF

A escolha deve ter em consideração duas dimensões: 1) suporte de um modelo estruturado de dados; e 2) garantia de eficácia e eficiência na devolução de dados.

Em [20], os autores relacionam as duas dimensões, mencionando a influência que uma exerce sobre a outra. O suporte do modelo de dados RDF, segundo os autores, deverá ter com conta os seguintes tópicos:

- Modelo abstrato de dados RDF: Qualquer modelo de dados representado em RDF deve ser representado como uma coleção de triplos, sendo que estes devem respeitar a estrutura representada na Figura 3:Triplo RDF. Este modelo de dados deve existir independentemente da sintaxe utilizada.
- Semântica Formal e Inferência: O RDF possui uma semântica formal que disponibiliza uma lógica na interpretação do grafo RDF. Segundo os autores, as regras de *reasoning* podem ser inferenciáveis como informação implícita através de informação explícita.
- Suporte para esquemas XML: Esquemas XML podem ser utilizados na representação de RDF. Estes esquemas XML, fornecem também uma estrutura que permite a extensibilidade para a definição de novos tipos de dados RDF. Contudo, esses novos tipos de dados devem ser suportados pela linguagem de consulta utilizada.
- Flexibilidade para a criação de novos *"statements"* para recursos: Em geral, devemos assumir que toda a informação não se encontra representada em RDF. Assim a linguagem de consulta deve tolerar a existência de informação incompleta e/ou contraditória.

Por outro lado, as linguagens de consulta devem respeitar o seguinte conjunto de propriedades[20]:

- Expressividade: A expressividade de uma linguagem de consulta normalmente indica o tipo de pesquisa que podemos fazer. Tipicamente, uma linguagem de consulta deve fornecer mecanismos semelhantes a álgebra relacional.
- Fechada: Com isto permite que o resultado dessa consulta possa ser parte integrante do grafo ou mesmo um novo grafo.
- Adequada: Uma linguagem de consulta é considerada adequada quando usa todos os conceitos do modelo de dados adjacente, ou seja, o resultado obtido de uma pesquisa não deve ser algo que não represente o modelo ou não possa ser interpretável do modelo.
- Ortogonalidade: Todas as operações podem ser utilizadas independentemente do contexto.
- Segura: A linguagem é considerada segura quando qualquer que seja a consulta devolve um resultado finito de dados.

3.1.1.1 Vocabulários RDF

À medida que foram sendo identificados termos específicos nos *statements* RDF, foram surgindo vocabulários[7], tais como, o FOAF[8], SIOC[9], SKOS[10], DCMI[11] e VOID[12] [22]. O vocabulário FOAF ,*The friend of a friend,* foi desenvolvido com o objetivo de descrever pessoas e as suas atividades. O SIOC, *The Semantic Interlinked online Communities,* é utilizado para a descrição de organizações e respetivas atividades. O DCMI, *Metadata Terms (DC Terms), é* usado com um propósito geral de relacionamento de meta-dados. O VOID, *Vocabulary of interLinked Datasets*, providencia meta-dados de um grande número de *Datasets*. O SKOS, *Simple Knowledge Organization System,* utilizado na representação de *thesaurus*[13], esquemas de classificação e taxonomias[14]. A utilização de novos vocabulários, veio ajudar na classificação e categorização de conteúdos da Internet. Contudo, tal dispersão de vocabulários, leva a problemas de integração e interoperabilidade dados entre plataformas.

3.1.2 OWL

Tal como o RDF, a linguagem OWL é utilizada no âmbito da *Web* Semântica, contudo direcionada para a publicação e partilha de ontologias[15] na Internet. A OWL, ou *Web Ontology Language,* carateriza-se pela expressividade na representação de conhecimento e compatibilidade com as linguagens XML da W3C. [23] distingue o RDF do OWL da seguinte forma: o RDF disponibiliza uma *framework* que permite estabelecer relações entre dados, enquanto que o OWL é melhor que o RDF na criação de restrições entre elementos de dados e suas relações. Atualmente, os formalismos de OWL mais utilizados são [25]:

- OWL Lite: Suporta necessidades de classificação hierárquica e restrições simples. Segundo [19], o OWL Lite é uma sub-linguagem do OWL DI (*description logic*) suportando apenas um subconjunto da linguagem OWL. Neste contexto, um subconjunto é definido com sendo um único vocabulário OWL.

---

[7] os vocabulários semânticos define um conjunto específico de termos (conceitos e relações) que poderão ser usados para a descrição/representação de uma visão particular de um determinado domínio.
[8] http://www.foaf-project.org/
[9] http://sioc-project.org/
[10] http://www.w3.org/2004/02/skos/
[11] http://dublincore.org/
[12] http://www.w3.org/TR/void/
[13] Lista de palavras com significados semelhantes, dentro de um domínio específico de conhecimento.
[14] Mecanismo de classificação hierárquica de conceitos
[15] http://www-ksl.stanford.edu/kst/what-is-an-ontology.html

- OWL DI (*description logic*): Suporta necessidades de classificação hierárquica e restrições simples. Por exemplo, não permite que um recurso seja usado como classe e instância.
- OWL Full: Possui maior capacidade expressiva, o que significa que consegue representar um maior número de aspetos da realidade do domínio em modelação, contudo o processamento computacional não é garantido, bem como pode não ser possível a obtenção de dados em tempo real. É considerada em [25], por combinar as especificidades do OWL DI com a liberdade sintática do RDF.

A linguagem OWL carateriza-se por ser uma linguagem de modelação com um propósito genérico tornando possível assim a sua utilização na representação de ontologias. Segundo [26], uma ontologia é um conjunto preciso de *"statements"* sobre um determinado domínio de interesse, representados por classes, propriedades, indivíduos e axiomas. As classes representam um conjunto de conceitos que possuem algo em comum. Em OWL, um membro de uma classe não é exclusivo, existe a possibilidade de disjunção de classes. As subclasses são usadas na representação de conjuntos equivalentes, não perdendo a relação hierárquica "pai e filho" entre a classe. As relações em OWL são designadas por propriedades, sendo que cada relação deve ligar: assuntos e domínios a objetos ou "conjuntos de objetos". De acordo com [26], estas propriedades podem ser classificadas em dois grupos: relações entre objetos e relações entre valores e objetos. As propriedades são importantes na construção do modelo, porque permitem a atribuição de um conjunto indeterminado de restrições, através da quantificação de objetos relacionados, ou por tipo de objetos. Existe também a possibilidade de atribuição de anotações no modelo OWL, anotações estas não utilizadas na lógica do modelo, mas que podem ser importantes para o utilizador na interpretação do mesmo.

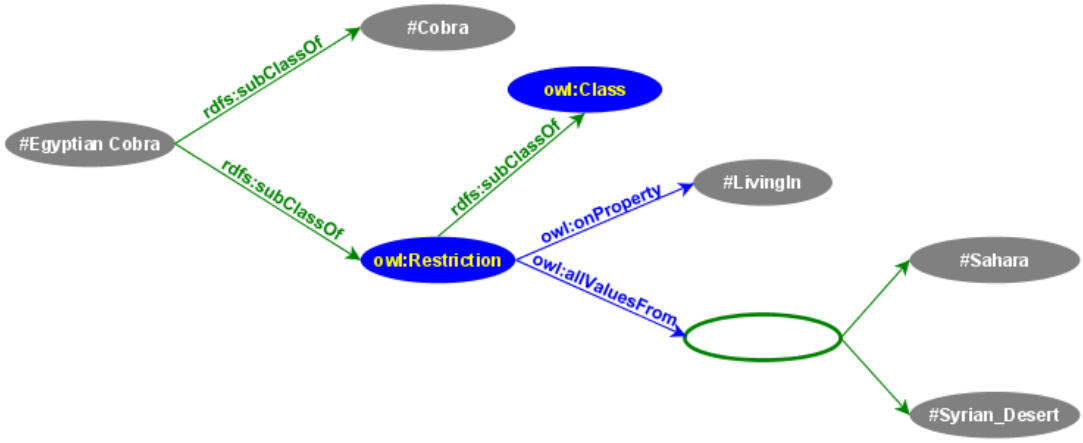


Figura 4: OWL[16]

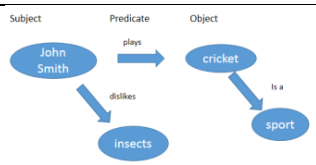
A Figura 4, corresponde a um exemplo de uma representação em OWL, da seguinte afirmação: "A cobra egípcia, é uma cobra que vive no sahara ou no deserto sirio".

3.1.3 Armazenamento de Dados em RDF

Os dados em RDF, podem ser armazenados em ficheiro, bases de dados nativas RDF ou ainda em bases de dados relacionais, através da sua customização. De um modo geral, a gestão de dados persistentes é efetuada através do recurso a bases de dados, controlando assim os problemas de concorrência e indexação de dados. Uma vez que, o armazenamento de dados em RDF é em forma de ficheiros, os problemas apresentados de concorrência e indexação permanecem. Pelo que as bases de dados nativas RDF, como Sesame[17], 3store[18], Redland[19], Jena[20], entre outros, são desenvolvidas de modo a suportar os triplos RDF, e, que normalmente disponibilizam o acesso a dados através de API's[21] e linguagens de consulta RDF. Segundo [16], as bases de dados relacionais (Ver tabela 2) podem suportar o armazenamento de modelos RDF através da criação de uma tabela com três atributos, que

[16] http://www.w3.org/2005/Talks/1111-Delhi-IH/#%2847%29
[17] http://www.openrdf.org/
[18] http://semantic*Web*.org/wiki/3store
[19] http://librdf.org/
[20] http://jena.apache.org
[21] Application Programming Interfaces

se identificam com os triplos RDF. Contudo, o autor, identifica alguns problemas de usabilidade deste tipo de implementação.



Object	Predicate	Subject
John Smith	Plays	Cricket
John Smith	dislikes	Insects
Cricket	Is a	Sport

Tabela 2: Representação de triplos em bases de dados relacionais

O controlo das consultas a dados, deveria ser com uma estrutura definida e controlada, de modo a evitar a devolução de dados inconsistentes, causando o "caos" na pesquisa de informação. Com a introdução do conceito *Linked Data* (ver secção seguinte), a dispersão das fontes, leva à necessidade de identificação dos triplos da sua fonte. Assim, [16] afirma que as *triple stores*[22] estão na verdade a evoluir para o que ele denomina de "*quad stores",* que adicionam um atributo que efetua a identificação da fonte.

3.2 *Linked Data*: conceitos base

Até ao momento abordamos o conceito da Internet semântica e a forma como este é materializado (em grande parte graças á tecnologia RDF), permitindo estruturar/representar a informação de modo a que esta possa ser facilmente partilhada e reutilizada entre aplicações, organizações e/ou comunidades. A Internet semântica possibilita, deste modo, que a informação esteja acessível usando uma arquitetura comum e genérica (Figura 5). A informação ganha uma nova dimensão, permitindo que esta se relacione entre si de uma forma *standard*. Contudo, e de acordo com a W3C[23], para que o conceito da Internet semântica seja efetivamente uma realidade, de modo a que indivíduos, organizações e comunidades possam usufruir dos benefícios ao nível da partilha e reutilização simplificada da informação, é necessário manter grandes volumes de informação disponíveis, acessíveis e geríveis. Ou seja, para além da possibilidade de descrever recursos e respetivas relações de forma comum (com RDF por exemplo), é necessário que esses mesmos recursos possam ser disponibilizados de forma igualmente partilhada. Este vasto conjunto de informação interrelacionada na Internet é denominado como ***Linked Data***. Adicionalmente ao RDF, o *Linked Data* pressupõe uma arquitetura de ou para integração em larga escala com capacidades de recuperação e inferência da informação disponível na rede.

[22] Base de dados específicas para armazenamento de dados em RDF
[23] http://www.w3.org/2001/sw/

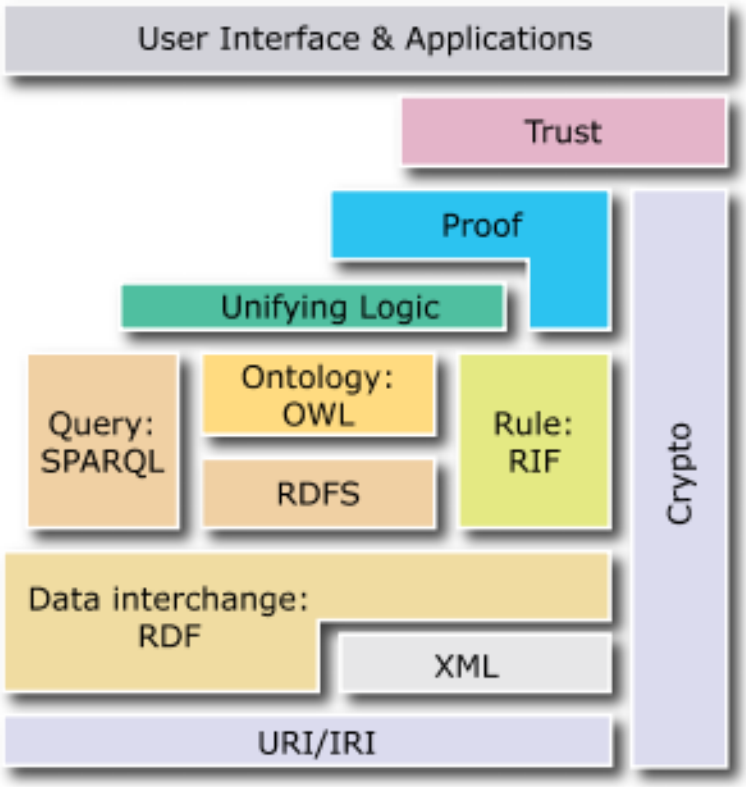


Figura 5:Arquitetura típica da *Web* semântica[24]

Numa visão mais pragmática, *Linked Data* é definido com sendo um conjunto de boas práticas usadas na publicação e interligação de dados na Internet [27]. Estas práticas encontram-se organizadas de acordo com um conjunto bem definido de quatro princípios:

Utilização de URI's[25] como identificadores;

Utilização de URI's HTTP, que referenciam conteúdos pesquisáveis;

Devolução de informação importante dos indentificadores URI's (por exemplo: RDF);

Inclusão de *links* URI para referenciação de documentos remotos.

Assim, a comunidade científica possui o interesse/desejo de incentivar a partilha de dados de forma estruturada e interoperável utilizando padrões da *Web* semântica.

Berners-Lee em [6], apresenta uma abordagem *bottom-up* de classificação para conteúdos a integrar na Internet semântica (ver tabela 3). Existe uma pirâmide de classificação do esquema de publicação de dados de uma a cinco estrelas, onde uma estrela

[24] http://www.w3.org/2007/03/layerCake-small.png
[25] Uniform Resource Identifier. Exemplo: http://www2.estgf.ipp.pt/cursos

representa dificuldade de integração segunda a abordagem *Linked Data*, e cinco ideais para integração segunda a abordagem *Linked Data*.

*	Publicação de dados não estruturados.
**	Publicação de dados estruturados.
***	Utilização de formatos não proprietários.
****	Utilização de URI's para identificação de conceitos/conteúdos.
*****	Ligação de conceitos próprios com descrições de terceiros.

Tabela 3:Classificação conteúdos Linked Data

A existência de um conjunto de boas práticas definidas para a comunicação e transferência de dados entre plataformas *Web*, conduziu a um crescimento exponencial de plataformas dotadas de capacidades de comunicação e partilha de informação. Assim, segundo [16], existe um conjunto de vantagens estudadas e identificadas que promovem a utilização do *Linked Data*, tais como:

Processamento/associação: Existência e/ou criação de *datasets* públicos que partilham um modelo uniforme de dados baseado em RDF. A utilização destes modelos, permite uma identificação mais clara e expressiva face a modelos de dados relacionais. Por exemplo, a definição de predicados[26] (entenda-se por elemento que efetua o relacionamento entre conceitos) dá ao utilizador uma clara ideia do tipo de relação existente entre conceitos. Estes modelos tornam a recuperação de informação uma tarefa mais simples com resultados de melhor qualidade.

Coerência: Um triplo RDF constituído por URI's de diferentes *namespaces* no *subject* e *object*, dá a entender que estabelece uma ligação entre a entidade que o identifica e o *subject* com a entidade que identifica o *object*.

Referenciação: Utilização de URI's para a referenciação e identificação de dados.

Integração: Desde que as fontes de dados partilhem modelos de dados RDF a integração semântica e sintática de *datasets* é facilitada por bibliotecas de integração como

[26] http://www.w3.org/TR/rdf-concepts/

por exemplo: RDF2GO[27]. Uma vez identificados os esquemas de dados, existem algoritmos de *matching* de dados entre os diferentes formatos.

Timeliness: publicação e atualização em *Linked Data* é relativamente fácil. Uma vez os dados estando dispersos e identificados por URI's as publicações passam pela disponibilização do URI à plataforma em questão. As atualizações devem ser cuidadas, mas uma vez alterada na fonte, é refletido em todos os locais onde esse URI estiver a ser utilizado.

No entanto, em [28], os autores apontam algumas questões de usabilidade, imprecisão, escalabilidade e dinamismo que consideramos importantes. No que diz respeito à usabilidade, os autores afirmam que existe a necessidade de ter conhecimento técnico da linguagem de consulta estruturada SPARQL[28]. O utilizador, para aceder aos dados, é obrigado a ter conhecimento da linguagem de consulta, bem como da estrutura de dados definida. Mesmo tendo em conta a existência de algumas *framework*s mais recentes, como RDF2GO, que procuram de certa forma ajudar o utilizador nas consultas, porém não deve ser de todo ignorada esta questão. A imprecisão, por sua vez, pode ter origem na grande diversidade de domínios que o *Linked Data* pretende englobar, mas a experiência dos utilizadores, bem como a utilização de *datasets* específicos de domínio ajudam na estruturação da informação. A escalabilidade e o dinamismo devem ser analisados como um todo, uma vez que, os conteúdos disponíveis na Internet são cada vez mais e diversificados. A constante mudança leva a que "o que hoje é verdade, amanhã já não o seja", então os domínios começam a crescer e os pontos de ligação são cada vez mais, o que conduz à necessidade de usar técnicas de *reasoning*[29] de ontologias [29]. Segundo [19], a ideia do *Linked Data* passa pela aplicação de uma arquitetura geral da *World Wide Web*, de modo a permitir a partilha da estrutura de dados à escala global. Com o objetivo de contribuir para este processo, dentro de alguns vocabulários (apresentados na secção seguinte), estão a ser implementados *endpoints*, ou seja, serviços que permitem o acesso a dados através da linguagem de consulta SPARQL. Da revisão da literatura verifica-se que existem três tipos de recolha de dados que correspondem à filosofia *Linked Data* [30]:

Tipo 1: Recorrendo única e simplesmente às capacidades do SPARQL e respetivos *endpoints*. Este mecanismo é bastante interessante quando é usado um SPARQL *endpoints*, ou um conjunto bem definido de SPARQL *endpoint* que permitam a catalogação de um conjunto de consultas bem definidas. Assim, é possível garantir com mais eficiência a

[27] http://semantic*Web*.org/wiki/RDF2Go
[28] http://www.w3.org/TR/rdf-sparql-query/
[29] Mecanismo de pesquisa que permite a realização de inferências baseado em factos e axiomas lógicos

qualidade e consistência dos dados. Contudo, a flexibilidade de adaptação a novos serviços requer um esforço superior e integração.

Tipo 2: Esta abordagem, consiste na criação de uma base de dados que efetue uma cópia integral de alguns *datasets* que se encontram definidos manualmente, de acordo com o domínio em questão. Desta forma, através de um único SPARQL *endpoint*, é disponibilizada toda a informação referente aos *datasets* anteriormente selecionados. Com esta abordagem é possível reduzir o número de consultas e ligações a efetuar, não sendo necessário um catálogo de consultas por SPARQL *endpoint*, como definido no tipo anterior. Contudo, o esforço de sincronização das cópias com os originais e a dificuldade na operacionalização, são bastante superiores.

Tipo 3: Utilização de um sistema agregador de consultas que permite a interligação de um vasto conjunto de *datasets*/SPARQL *endpoints*. Este sistema permite que seja efetuado um pedido de consulta ao mediador, sendo o pedido direcionado com total "transparência" para o emissor do pedido. O mediador "faculta" dados íntegros, no entanto, o tempo de resposta é superior e existe uma necessidade constante de manter o mediador atualizado e funcional, uma vez que este tipo ao contrário do Tipo 2, não possui nenhum mecanismo de *cache* que permita um tempo de resposta mais curto e também uma resposta rápida da *cache* quando o acesso ao original se encontra congestionado.

Tal como apresentado na literatura de *Linked Data* e apresentado em [30], o tipo um e três são aqueles aos quais é atribuída melhor confiança e fiabilidade. Assim, no decorrer deste trabalho, todo o foco, vai ser sobre o tipo um e três, como podemos verificar posteriormente no capítulo 4.

Tendo em conta estes princípios *Linked Data*, a comunidade científica começou a detetar algumas vantagens da sua utilização na integração de informação entre plataformas empresariais, que são apresentadas na subcapítulo seguinte.

3.3 Plataformas/*Frameworks* de integração baseadas em *Linked Data*

Na base da motivação do *Linked Data*, encontra-se a problemática da integração, em particular integração em rede e em larga escala. E, é um facto da literatura que, a integração de informação é uma problemática muito estudada no contexto de sistemas de informação. Segundo [31], a integração de dados em grandes empresas é crucial, com custos elevados, bem como é caraterizado por ser um processo demorado e problemático. A integração de

informação entre sistemas ERP, CRM e SCM que atualmente são os sistemas mais utilizados pelas empresas bem como as suas grandes fontes de informação, é um grande desafio. De forma a colmatar algumas das falhas de comunicação entre estes sistemas, algumas abordagens que utilizavam XML, como *Web services* e SOA.

Em termos genéricos, identificam-se cinco indicadores ou abordagens a ter em conta no desenvolvimento de arquiteturas de integração de plataformas empresariais [31]:

- **Indicador 1 –** Abordagens baseadas em taxonomias: As taxonomias permitem a disponibilização de um modelo linguístico que permite classificar grandes quantidades de documentos, emails, diretivas, etc, no fundo todo o tipo de dados utilizados no dia-a-dia. Algumas ferramentas existentes no mercado, como o Microsoft sharepoint[30] e termStore[31] permitem a construção e utilização de taxonomias, contudo a hierarquia rígida e duplicação de conceitos é um problema presente.
- **Indicador 2 -** Utilização de esquemas em XML: A utilização de esquemas em XML nas empresas é uma prática comum desde o seu aparecimento em 1998, contudo desde então já são conhecidos e identificados cerca de 54960 esquemas distintos. Cada organização desenvolve o seu próprio esquema o que leva assim à necessidade de criação de mecanismos de mapeamento em XML, conduzindo ao crescimento de custos destes mecanismos de mapeamento
- **Indicador 3 –** Abordagens baseadas na utilização de wikis: As wikis encontraram-se em grande crescimento de utilização nos últimos anos, o seu crescimento teve em conta os requisitos das empresas, como controlos de acessos, segurança, entre outros. Assim podemos dizer que existem wikis para todos os gostos, dependendo dos cenários, propósito e se lida ou não com documentos. Surgiram alguns wikis como: Confluence[32], Jive[33], Twiki[34]. O problema das wikis, é que torna a informação percetível para o utilizador mas não para máquinas
- **Indicador 4 -** Integração com bases de dados relacionais: As bases de dados relacionais, encontram-se bem implementadas e bastante utilizadas no contexto de sistemas de informação. Para criar uma vista que unifique um conjunto de bases de dados diferentes, podem ser utilizados diferentes métodos, como por exemplo

[30] http://office.microsoft.com/pt-pt/sharepoint/
[31] http://office.microsoft.com/en-us/office365-sharepoint-online-enterprise-help/manage-the-term-store-and-managed-metadata-HA102476167.aspx
[32] https://www.atlassian.com/software/confluence
[33] http://br.jivesoftware.com/
[34] http://twiki.org/

agregadores de *queries*. Os agregadores de *queries* podem ser utilizados quando falamos de tecnologias compatíveis e semelhantes. O caso muda de figura, quando possuímos fontes de dados heterogéneas e dispersas. Esta integração heterogénea tem um custo elevado para a empresa [31].

- **Indicador 5 –** Abordagens baseadas na centralização da autenticação: Sistemas robustos e centralizados de gestão de utilizadores, que efetuam um controlo de permissões de acesso a conteúdos.

Face aos indicadores aqui apresentados, a tabela 4 mostra de quais e de que forma os princípios *Linked Data,* podem ajudar na integração de informação em plataformas empresariais.

Indicadores	Principios *Linked Data de apoio aos indicadores*
1	• Todos os termos se encontram associados a um identificador único (URI), assim todos os conceitos podem ser acedidos através de um *browser Web*, não necessitando de *software* adicional. • A responsabilidade pela manutenção da estrutura de termos, pode ser repartida por diversos departamentos distribuindo assim tarefas e responsabilidades. • Utilizando o vocabulário SKOS, os termos podem ser agrupados de forma hierárquica e os problemas de granulidade, ou seja, de duplicação de conceitos, são resolvidos. • Os dados podem ser representados numa estrutura RDF. • Podem ser atribuídos múltiplos conceitos que representam literais RDF, ou seja, conceitos chave. • Finalmente, o resultado de uma implementação de *Linked Data* pode ser re-utilizado para futuras implementações em diferentes contextos.
2	• Introduz a necessidade da aplicação de um conjunto limitado de esquemas para domínios comuns. • A utilização de ferramentas como o ontowiki, e SPARQL *queries*, permite aos utilizadores efetuar alterações e pesquisas nas estruturas de dados de forma mais fléxivel, face aos sistemas de XML dispersos e rígidos controlados pelos departamentos de *Corporate Language Management* (CLM) das empresas. • O principal desafio passa pela mudança onde o mapeamento do XML com esquemas RDF é essencial..
3	• Contribuiu para o desenvolvimento de wikis, mais elaboradas e com capacidades semânticas, como a Semantic Media Wiki[35]. • A nível empresarial a utilização de wikis como Ontowiki trazem vantagens como: o foco na informação estruturada torna-a disponível a ser utilizada por outras aplicações; toda a

[35] http://semantic-mediawiki.org/

	informação se encontra automaticamente possível de disponibilizar de acordo com os princípios de *Linked Data*, pois toda se encontra mapeada em RDF; a informação fragmentada pode ser ligada com recursos da empresa; a informação textual pode ser representada em RDF através da sua classificação.
4	• Permite a identificação de todos os recursos com o URI, onde podemos identificar recursos e/ou bases de dados relacionais com bases de dados não relacionais. • Todas as pesquisas são efetuadas recorrendo à linguagem de pesquisas SPARQL. • O desafio passa essencialmente pelo mapeamento entre *triple stores* e bases de dados relacionais de forma a reduzir a duplicação de dados bem como, a uniformização e desenvolvimento em conjunto com linguagens de pesquisa robustas.
5	• Mais sofisticada não recorre a *passwords*, mas sim a perfis de utilizadores, essencialmente suportado pelo vocabulário FOAF

Tabela 4:Principios Linked Data de apoio aos indicadores

Considerando os cinco identificadores identificados anteriormente, foi efetuada uma pesquisa de plataformas e /ou *framework*s que:

- **Critério 1 -** Efetuassem pesquisas semânticas, que permitam a criação de pesquisas eficientes onde possam ser implementadas pesquisas facetadas, ou navegação pelos resultados;
- **Critério 2 -**Devolvessem resultados semânticos relevantes que ajudem na identificação da informação;
- **Critério 3 -** Possuíssem capacidades semânticas na integração de informação, através da recuperação de informação de diversas fontes de informação;
- **Critério 4 -** Permitissem a personalização e customização para um domínio específico;
- **Critério 5 -**Permitissem uma interoperabilidade entre vocabulários (ex: SIOC, FOAF, RDF), bem como comunicação entre diversas plataformas.

Enquadrado pelos cinco desafios a ter em conta no desenvolvimento de arquiteturas de integração de plataformas empresariais e pelas condições definidas nos pontos imediatamente anteriores, apresenta-se na seção seguinte uma seleção e análise de algumas plataformas/*frameworks* relevantes para o contexto do presente trabalho.

3.4 Abordagens à integração de informação empresarial

Para a integração de informação entre empresas, recorrendo aos princípios *Linked Data*, foram analisadas duas abordagens distintas à resolução do problema. Assim, as abordagens orientadas a wikis e conteúdos/documentos, foram analisadas.

3.4.1 Wikis

As abordagens orientadas a wikis utilizadas para a integração de informação entre sistemas empresariais, seguem os seguintes princípios [32]:

- **Wikis permitem qualquer pessoa editar:** Em wikis, o conceito de restrições e hierarquias restritas é eliminado, dando origem a uma filosofia onde qualquer um pode dar o seu contributo e expressar a suas ideias;
- **Wikis são fáceis de utilizar:** Qualquer utilizador deve sentir-se familiarizado com as funcionalidades de processamento de texto (escrita, remoção e armazenamento);
- **Todos os conteúdos são redirecionáveis (*linkable):*** os utilizadores podem criar *links* entre palavras e/ou estabelecendo relações semânticas;
- **Wikis permitem gestão de versões:** A informação nunca desaparece da wiki. A informação pode ser editada, removida ou adicionada, mas encontra-se sempre no controlo de versões da wiki;
- **Wikis suportam todo o tipo de conteúdos.**

Como enunciado no subcapítulo anterior, podemos dizer que existem wikis para todos os gostos, acrescentando e para todas as linguagens de programação. De entre uma lista extensa de wikis : TiddlyWiki[36], MediaWiki[37], Ontowiki, Kiwi[38], Zwiki[39], Twiki, Semantic Media Wiki entre outros. Num contexto mais orientado à integração de informação de sistemas empresariais, as plataformas ontowiki e kiwi, são aquelas que mais aparecem referenciadas na literatura.

A plataforma ontowiki aparece como uma ferramenta ágil de apoio à gestão de conhecimento distribuído. É uma ferramenta desenvolvida em PHP, capaz de utilizar bases

[36] http://tiddlywiki.com/
[37] http://www.mediawiki.org/wiki/MediaWiki
[38] http://kiwi-project.eu/
[39] http://zwiki.org/FrontPage

de dados estruturadas de suporte como o MySQL[40] ou até *triple stores* como o Virtuoso. Esta é uma ferramenta *Web*, desenvolvida com especial preocupação para o *user interface* onde é possível visualizar mecanismos de suporte a pesquisas facetadas, mecanismo de controlo de versões e áreas colaborativas, que permitem a criação de tópicos de discussão sobre determinados conceitos. Esta é uma plataforma com base na Powl [33], para a qual foram definidas as seguintes funcionalidades [34]:

- Interface de apresentação e edição de dados intuitivo, independentemente do domínio onde é aplicado;
- Vistas semânticas que permitem a geração e agregação da base de conhecimento;
- Evolução e controlo de versões, disponibiliza a oportunidade de controlar a evolução do conteúdo, bem como de retroceder, corrigindo problemas;
- Pesquisa semântica fácil de utilizar, permitindo pesquisa livre e criação de filtros nos resultados;
- Comunidade de apoio, onde a discussão possui espaço privilegiado. Os utilizadores são "convidados" a votar sobre determinados assuntos bem como a dar as suas opiniões;
- Estatísticas, permitem interatividade na medida de popularidade dos conteúdos através da atividade dos utilizadores;
- Permite a distribuição semântica da informação, de modo a que a informação possa ser utilizada por outras aplicações.

Com o objetivo de dar resposta aos requisitos funcionais agora listados, foram implementados mecanismos que:

- Disponibilizam interfaces "amigáveis" em AJAX, que permitem uma navegação de forma mais fluída;
- Permitam a criação de vistas, e combinação de vistas, disponibilizando pesquisas facetadas e livres, integração com API de Google Maps e vistas que utilizam calendários;
- Disponibilizam uma componente social, suportada por registo de alterações, possibilidade de utilização de comentários, atribuição de *ratings*, estatísticas que permitem consultar a popularidade de acesso à informação.

O projeto Kiwi, apresentado em [32], onde foi estudada uma abordagem que deu origem a uma plataforma com características sociais com princípios de wiki, segue uma

[40] http://www.mysql.com/

abordagem denominada *"Content Versatility"*. Esta abordagem pretende efetuar uma combinação de conteúdo *human-readable* e metadados, onde o mesmo conteúdo pode ser apresentado em diferentes locais como wikis, *blog* entre outros.

3.4.2 CMS- Content Management Systems

Na vertente orientada a CMS e a conteúdos em geral, surgem algumas abordagens como a *Linked Data Content Management System* [35]. Nas abordagens de gestão de conteúdos tradicionais, onde são utilizadas bases de dados relacionais e existem funcionalidades de edição, adição e remoção de conteúdos, onde os utilizadores já se encontram familiarizados. Agora os conteúdos registados nos sistemas tradicionais com bases de dados relacionais, necessitam de ser coordenados com os princípios do *Linked Data*. Esta coordenação, passa pelo mapeamento dos conteúdos existentes nas bases de dados relacionais, com as *triple stores*, quer através da atribuição de identificadores únicos (URI's) aos conteúdos, quer através de ontologias para classificação de conteúdos de um determinado domínio. Neste sentido são utilizadas pesquisas estruturadas, normalmente às classes da ontologia em SPARQL.

A plataforma PoolParty[36] , surge essencialmente pela necessidade de uma ferramenta de fácil utilização para a gestão de *thesaurus* para a *semantic Web.* O interface desenvolvido em AJAX, disponibiliza aos utilizadores a capacidade de importação de *thesaurus* que podem estar esquematizados em diferentes tipos de serializações em RDF (RDF/XML, N-Triples ou Turtle), respeitando sempre o vocabulário SKOS. O interesse despertado por alguns dos participantes do projeto levou à continuação do mesmo. Atualmente a ferramenta PoolParty é constituída por cincos componentes: *PoolParty Thesaurus Server (PPT), PoolParty Extractor (PPX), PoolParty Semantic Search (PPS), PoolParty PowerTagging (PPP)* e *PoolParty Semantic Integrator (PPI).* No contexto deste trabalho, são analisados três dos cinco componentes.

O componente PPT, permite uma gestão e controlo de vocabulários, como taxonomias ou *thesaurus*, baseados em SKOS e/ou RDF. Neste componente é também disponibilizada uma API que permite a sua integração com outros sistemas, como CMS.

O PPI, permite as empresas consolidar coleções de dados de diversas fontes e heterogéneas. Seguindo os princípios do *Linked Data* toda a informação tem que ser representada num grafo de conhecimento. Assim, permite efetuar pesquisas avançadas em SPARQL recorrendo à cláusula SPARQL *ASK.*

PPS disponibiliza uma API que permite pesquisa por texto livre, pesquisa facetada, pesquisa *auto-complete*, pesquisa por similaridade.

Outra plataforma analisada foi a *Linked Media Framework* (LMF) que teve como base para o seu desenvolvimento o projeto Kiwi [37], é constituída por cinco módulos distintos. Esta plataforma surge pela necessidade de colmatar três problemas detetados em [37]:

- Como estender os princípios do *Linked Data*, através do acréscimo de serviços REST-Ful na adição, modificação e remoção de recursos;
- Como estender os princípios de *Linked Data,* na gestão de conteúdos e metadados usando a norma *Multipurpose Internet Mail Extensions* (MIME) [41] e mapeamento URL;
- Como oferecer um caminho que permita pesquisas em recursos *Linked Data* de forma mais simples.

Para dar respostas às necessidades apresentadas para o desenvolvimento da *framework* LMF, foram desenvolvidos dos seguintes módulos:

- LMF Core implementa um servidor *Linked Data*, bem como os serviços propostos em [37]. Os serviços implementados, denominados por *ResourceWebService* são implementados de modo a trabalhar sobre o servidor *Linked Data*, responsável pela gestão de triplos, transações e gestão de versões, em que está preparado para armazenar a proveniência da informação;
- LMF *Search*: módulo que oferece a capacidade de pesquisa semântica através do recurso ao Apache Solr[42]. Este módulo é altamente customizável, pelo facto de permitir o mapeamento de conjuntos de regras para RDF, SKOS, Dublin Core. Este módulo pode ser acedido através de um REST API;
- LMF SPARQL: que disponibiliza um SPARQL *Endpoint* para a pesquisa e atualização de dados obtidos na instalação. Este módulo é suportado pelo SESAME[43] e SPARQL 1.1.
- LMF *Linked Data Caching* disponibiliza um mecanismo de *cache* transparente utilizável pelos utilizadores em SPARQL.
- LMF *Reasoner:* que oferece um conjunto de regras baseadas em sKWRL [38], desenvolvido no projeto Kiwi.

Com resultado da análise do estado da arte deste trabalho, de entre muitas plataformas analisas, surgiram as quatro anteriormente apresentadas, pois foram aquelas que

[41]

[42]http://lucene.apache.org/solr/

[43] http://www.openrdf.org/

melhor se enquadravam no contexto do mesmo. Assim, foi possível chegar às seguintes conclusões:

- Apesar de terem sido analisadas duas plataformas para cada uma das abordagens apresentadas (wiki e orientada aos conteúdos), a preferência recai sobre as plataformas orientadas a conteúdos.
- É também privilegiada uma escolha sobre *framework*s/serviços, na medida em que este trabalho passa pela integração de dados entre plataformas, dito por outras palavras, é pretendido enriquecer as plataformas com informação relevante e direcionada;

Serviços/*frameworks* que respeitem princípios do *Linked Data.* Este é um requisito fundamental, na medida em que é pretendido o recurso a um conjunto de boas práticas que se encontra em constante evolução, atribuindo garantias de utilização futura.

3.5 Síntese das plataformas/*frameworks*

Na tabela 4 é apresentada uma comparação das plataformas anteriormente apresentadas, tendo em conta os seguintes critérios:

- **Usabilidade:** caracterizada pela disponibilização de mecanismos de pesquisa facetada e acesso único (através de *browser*) e manipulação de dados em *real-time*(Ajax);
- **Customização:** disponibilização de vistas Semânticas (Genéricas ou Especificas);
- **Generalização:** suporte de utilização e manipulação de múltiplas ontologias;
- **Colaboração:** controlo de acessos, suporte de vocabulários RDF e RDFa, componente social suportada por comentário e *ratings*;
- **Portabilidade:** Suportado pela grande maioria de *browsers*;
- **Pro-atividade**: Reutiliza conceito e efetua sugestões;
- **Interoperabilidade:** Suporte de formatos *standard* RDF;
- **Escalabilidade.** Suporta cache, estratégia definida de armazenamento (Mysql, Virtuoso).

Critérios	Plataforma Orientada			
	Wikis		Gestão de conteúdos	
	Ontowiki	Kiwi	LMF	PoolParty
Usabilidade	Possui um interface *Web*, baseado no conceito RIA[44] em AJAX. Implementa um mecanismo de pesquisa faceta a ontologias.	Possui um interface *Web* dinâmico, (*wiki based*), com suporte a mecanismos de pesquisa facetada	*Framework*, com interface *Web* de administração. Serviço disponibilizado através do protocolo de comunicação RESTFul.	Ferramenta com interface *Web* RIA, para a manipulação de taxonomias representadas em SKOS. Suporta pesquisas facetadas
Customização	Recorre RDF *search engine*, Ontogator[45], na implementação de vistas semânticas	Plataforma "aberta", com vistas executadas tendo em conta o perfil de utilizador.	Permite a restrição a dados, através do interface de administração	Permite a pesquisa e manipulação de taxonomias individualmente.
Generalização	Permite pesquisa e manipulação de múltiplas ontologias	Permite a pesquisa em múltiplas ontologias	Permite manipulação e pesquisa em múltiplas ontologias	Permite manipulação e pesquisa em múltiplas taxonomias
Colaboração	Componente social suportada por um mecanismo de *rating* e comentários e controlo de versões.	Mecanismo de colaboração suportado por um mecanismo de comentários		Componente social suportada por um mecanismo de comentários.
Portabilidade	Suportado pela maioria dos *browsers*	Suportado pela maioria dos *browsers*	Mecanismo só com interação com o cliente através de um protocolo	Suportado pela maioria dos *browsers*

[44] Rich Internet Application
[45] http://www.seco.tkk.fi/projects/sem*Web*/dist.php

			de comunicação RESTFull	
Pro-Atividade	Permite a reutilização de conceitos na construção da ontologia.	Sistemas reativos. Responde às pesquisas efetuadas	Pode ser utilizado como mecanismo de sugestões através de *inputs* bem definidos	Permite a reutilização de conceitos em taxonomias SKOS
Interoperabilidade	Suporta os diversos vocabulários RDF.	Suporta os diversos vocabulários RDF	Suporta os diversos vocabulários RDF	Suporta o vocabulário RDF, SKOS.
Escalabilidade	Permite a integração com base de dados MySql e Virtuoso	Utiliza uma gestão de dados interna própria.	Interage com qualquer base de dados através de SPARQL *endpoints*	Utiliza a base de dados Sesame, aplicando os mecanismos de gestão de *cache* da mesma

Tabela 5:Sintese de plataformas/frameworks

A tabela 5, apresenta os oito critérios que foram selecionados na análise das quatro plataformas, onde na última coluna é efetuada uma especificação mais pormenorizada das caraterísticas relevantes para este trabalho.

Com conclusão deste capítulo, verificou-se que a plataforma a plataforma ontowiki é aquela que implementa um maior número de funcionalidades que vai de encontro aos requisitos definidos. Contudo, dentro das plataformas wiki, foi detetada uma caraterística na componente social, nomeadamente no que diz respeito à gestão de permissões, de utilizadores, onde as wikis são demasiado limitadas. Face a este resultado, e uma vez que falamos em integração de dados em plataformas empresariais, onde a informação e respetivas permissões é um tema bastante delicado. Levou a que a abordagem a seguir a ser baseada em CMS.

Capítulo 4

Definição da solução

4.1 Contextualização

A solução apresentada nesta dissertação, surge no contexto de uma necessidade real de um projeto europeu, denominado H-Know[46]. Este capítulo apresenta o desenho de uma solução de engenharia que possibilita a partilha e reutilização de informação entre plataformas colaborativas de dados estruturados. Esta solução, vai ao encontro do estado da arte apresentado no capítulo dois e três, na medida em que implementa uma arquitetura, que recorre a tecnologias e princípios do *Linked Data*.

O projeto H-Know foi um projeto financiado pela comissão europeia (refª NMP-2007-214567) entre 2009 a 2011, contando com 15 parceiros de 5 países diferentes. A plataforma tecnológica H-Know, trata-se de uma plataforma colaborativa de dados estruturados que, de acordo com os objetivos citados na página do projeto, pode ser apresentada da seguinte forma: "solução oferecida às pequenas e médias empresas da área da restauração e manutenção de edifícios antigos, para aceder a conhecimento específico partilhado por uma comunidade. As pequenas e médias empresas, podem partilhar conhecimento, potenciando a criação de parcerias e colaborações".

4.2 Apresentação dos requisitos

Tendo o trabalho sido enquadrado num determinado domínio, apresentado no capítulo dois e três desta tese, onde o principal foco passa pela integração de informação entre

[46] http://www.h-know.eu

plataformas colaborativas de domínio empresarial recorrendo aos princípios *Linked Data,* surge então a necessidade de efetuar um levantamento de requisitos.

Segundo [39], o levantamento de requisitos no domínio de sistemas de informação é um processo moroso e difícil, que necessita de ser executado por profissionais que na maioria das vezes, privilegiem o contato com os *stakeholders*/interessados. Neste trabalho, o levantamento de requisitos, foi uma tarefa desenvolvida recorrendo aos parceiros do projeto H-Know. Uma vez, que o resultado da solução proposta será algo com interesse para os mesmos, o processo de comunicação e colaboração no levantamento de requisitos, foi bastante ágil. Contudo em [39], são apresentadas algumas técnicas que ajudam neste processo de levantamento de requisitos em casos mais complexos.

De acordo com a revisão da literatura, análise de requisitos deve ser direcionada no sentido do levantamento de requisitos funcionais e requisitos não funcionais.

A interação com os parceiros do projeto H-Know para o levantamento de requisitos funcionais, deu origem á seguinte lista de requisitos:

- O utilizador poderá efetuar uma pesquisa de texto livre a um número diversificado de fontes orientado a um domínio específico;
- O utilizador poderá definir filtros de pesquisa, por exemplo, pesquisa local, *cache* do sistema, ou todas a fontes disponíveis;
- O utilizador deverá ter acesso a uma ontologia que ajuda no refinamento das suas pesquisas;
- Deverá devolver conteúdos de sugestão, para espaços colaborativos, com classificação atribuída;
- Deverá existir dois tipos de perfis de utilizadores: administrador e cliente.

De modo a ajudar na formalização dos requisitos funcionais, foi elaborado um diagrama de casos de uso apresentado na Figura 6.

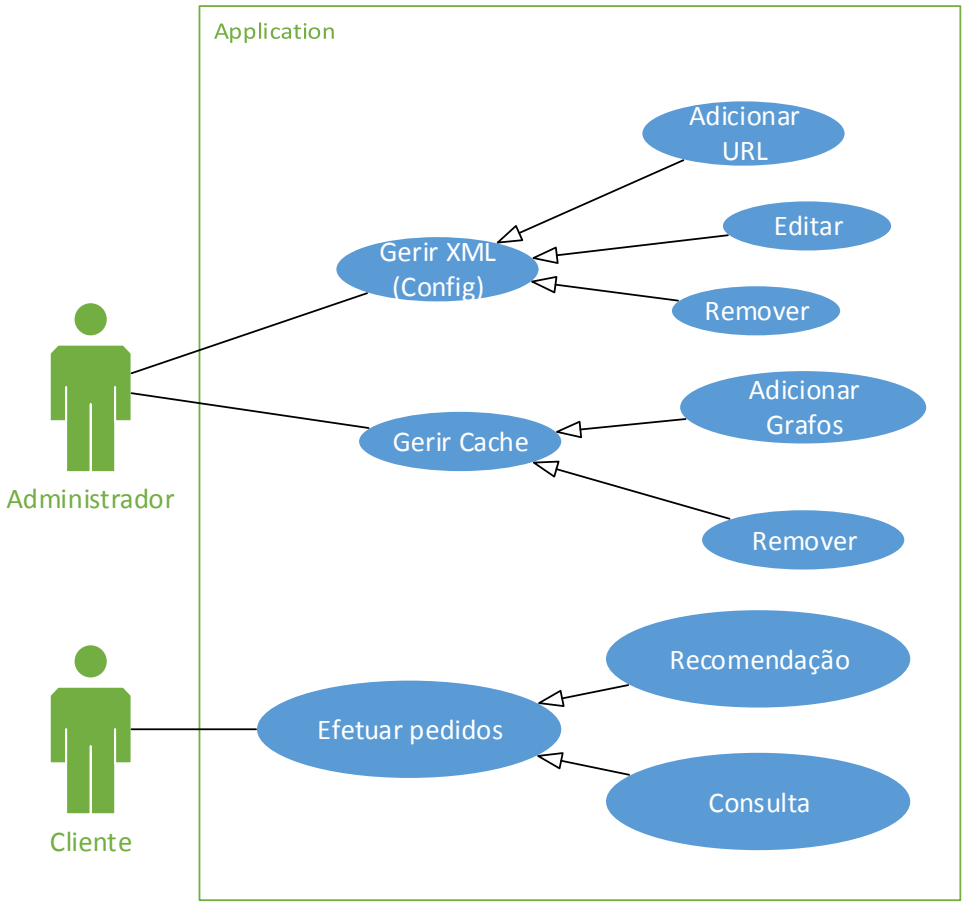


Figura 6:Casos de uso (Requisitos Funcionais)

De acordo com o diagrama apresentado na Figura 6, consideramos dois atores:

- Administrador: Responsável tecnológico, conhecedor do serviço, que vai interagir com o ficheiro de configuração XML. Este ficheiro de configuração possui os *SPARQL endpoints* que irão ser utilizados nas consultas pelo serviço;
- Cliente: O cliente será uma plataforma de dados estruturados que deverá ser capaz de gerar e enviar um ficheiro XML, que contenha a estrutura adequada à leitura no serviço. Este cliente deverá também ser capaz de processar/interpretar a resposta em XML resultante do pedido.

Os requisitos não funcionais, foram efetuados numa segunda iteração, através de uma análise ao estado da arte, onde foi analisado estado atual do projeto H-Know, bem como, uma revisão mais generalizada do estado da arte onde foram analisadas plataformas semelhantes, boas práticas e princípios. Como resultado dessa pesquisa que se encontra fundamentada no estado da arte desta dissertação, destacaram-se os seguintes requisitos não funcionais.

A solução a desenvolver deverá:

- ser um serviço desenvolvido recorrendo ao protocolo de comunicação RESTFul;
- utilizar tecnologias, vocabulários e esquemas de informação que se encontrem referenciados nos princípios do *Linked Data;*
- ter um interface para o utilizador de fácil utilização nas pesquisas de conteúdos.

De notar que os requisitos funcionais e não-funcionais, se encontram alinhados com os indicadores a considerar no desenvolvimento de arquiteturas de integração de plataformas empresariais, firmados no capítulo 3 – subcapítulo 3.3, fundamentalmente os que indiciam preocupação semântica, ou seja, o indicador 1 e o indicador 2. Prevê-se a implementação dos requisitos garantindo o aplicação dos princípios *Linked Data* que estendem os indicadores mencionados, colmatando algumas das suas restrições tal como sistematizado na tabela 4 do subcapítulo 3.3. Adicionalmente, os requisitos funcionais e não funcionais procuram visar os critérios base (capítulo 3 – subcapítulo 3.3) que definem as preocupações de pesquisa e integração semânticas a suportar numa plataforma colaborativa empresarial.

Segundo as boas práticas da engenharia de *software*, o levantamento de requisitos deve estar associado ao desenvolvimento de um documento formal de requisitos. A construção deste documento pode ser efetuada segundo diversos *standards* já desenvolvidos e que são muitas as vezes constituintes da documentação contratual de um projeto[40].

No contexto deste trabalho o levantamento de requisitos, foi essencial para o enquadramento do trabalho num determinado domínio, escolhas de tecnologias e desenho da arquitetura. Nesta fase, o desenho da arquitetura começa a ficar mais clara como também a necessidade uma solução que permita:

- ao cliente/aplicação interagir, recorrendo ao protocolo de comunicação RESTFul;
- aceder através de um interface amigável de pesquisa a fontes de conteúdos dispersas, com o auxilio de uma ontologia na aplicação de filtros de refinamento da pesquisa;
- garantir um mecanismo de *cache* interna, com fontes geridas pelos administrador e resultantes de pesquisas de utilizadores;
- ao administrador uma gestão de fontes de informação fácil.

Uma vez obtidos e validados os requisitos, o passo seguinte será efetuar um levantamento de tecnologias que possam suportar o desenvolvimento desta solução e que serão apresentadas no subcapítulos seguinte.

4.3 Tecnologias em *Linked Data*

No seguimento da apresentação dos requisitos e considerando que a sua implementação pressupõe a aplicação ou uso dos princípios e tecnologias *Linked Data,* revela-se pertinente caracterizar com um pouco mais de detalhe a tecnologia que incorporará a proposta de solução descrita no presente capítulo. Deste modo, apresentam-se, sob uma perspetiva mais técnica: duas *frameworks,* duas bases de dados, um servidor universal[47] e um servidor de consultas.

4.3.1 *Framework* Jena

Jena é uma API Java que permite aos programadores não familiarizados com esquemas RDF, mas sim com um conjunto de métodos e objetos em Java[48]. Esta API possui um processador de consultas (ARQ), que permite a sua utilização no processamento de consultas com recurso a linguagem SPARQL, através de SPARQL *endpoints*.

4.3.2 *Framework* Sesame

O sesame[49] é uma ferramenta *open source*, composta por uma *framework* de manipulação de RDF, por uma base de dados RDF e ainda mecanismos que suportam a inferência e *queries* a dados. Esta uma é uma ferramenta com uma API para programadores em JAVA e pode ser utilizada com bases de dados RDF que não seja a sesame, como por exemplo : 4store[50], BigData[51].

[47] Ferramenta que agrega um diverso conjunto de funcionalidades *Linked Data* (armazenamento, processamento, disponibilização de dados)
[48] http://jena.apache.org/tutorials/rdf_api.html
[49] http://www.w3.org/2001/sw/wiki/Sesame
[50] http://www.w3.org/2001/sw/wiki/4store
[51] http://www.w3.org/2001/sw/wiki/Bigdata

4.3.3 Servidor de dados Virtuoso

O Virtuoso[52] é um servidor de dados "multimodelo", que segundo a OpenLink[53] (empresa que o desenvolve) pode ser utilizado em empresas ágeis. Este servidor é disponibilizado pela empresa através de uma versão *opensource*, que pode ser utilizada pela comunidade, focando-se o seu modelo de negócio na prestação de serviços. A Figura 7 apresenta a arquitetura detalhada do Virtuoso. O Virtuoso é um servidor bastante completo, pois é o único servidor *opensource* que possui um conjunto tão diversificado de serviços direcionados ao *Linked Data.* A utilização de *standards* e protocolos de comunicação facilita a interoperabilidade entre serviços, bem como a compreensão dos programadores na sua utilização.

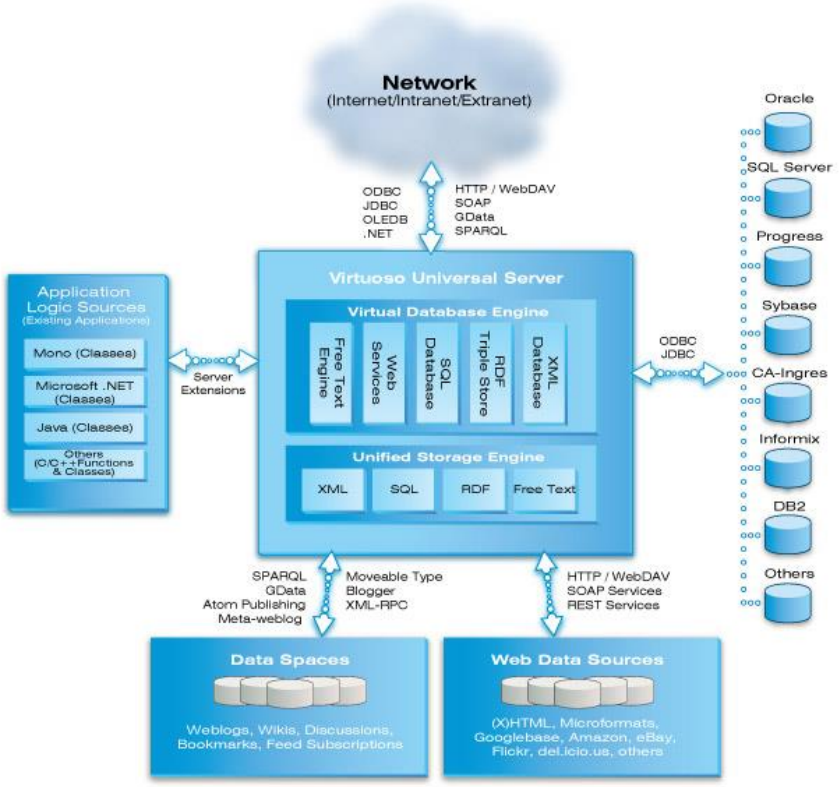


Figura 7:Arquitetura Virtuoso[54]

[52] http://virtuoso.openlinksw.com/
[53] http://www.openlinksw.com/
[54] http://virtuoso.openlinksw.com/

4.3.3.1 *Middleware* Virtuoso Sponger

O virtuoso sponger[55] é um componente *middleware* do Virtuoso *universal server* e funciona como agregador de fontes de dados em diversos formatos, disponibilizando a informação de forma transparente e integrada no processador de *queries* do Virtuoso *universal server.*

4.3.4 Servidor Apache Solr

O Apache Solr[56] é um servidor de consultas que recorre a REST-API[57] para efetuar a indexação da localização de ficheiros em XML, JSON[58], CSV[59] ou de forma binária através de HTTP. [41] enuncia as seguintes características para este servidor:

- Capaz de efetuar consultas avançadas em todo o texto;
- Otimizar pesquisas em grandes volumes de dados;
- Possuir interfaces *standards* de comunicação, autorreplicação e recuperação em caso de falhas;
- Indexação em tempo real e configuração flexível através de XML.

4.4 Arquitetura da solução

Uma vez apresentadas as tecnologias mais referenciadas na literatura do domínio de *Linked Data,* é apresentada na Figura 8, a arquitetura da solução sob a forma de um diagrama de componentes. A arquitetura é composta por três componentes, uma aplicação e uma referência á *Web* através onde se encontram SPARQL *endpoints* e outras fontes de dados estruturados.

Na Figura 8, destaca-se o componente "*Semantic Search Service*" correspondendo ao core da solução de engenharia desenhada neste subcapítulo. De um modo geral a Figura 8, pretende exemplificar qual e quais os tipos de comunicações que podem ser efetuados pela

[55] http://virtuoso.openlinksw.com/dataspace/doc/dav/wiki/Main/VirtSponger
[56] http://lucene.apache.org/solr/
[57] http://www.ibm.com/developerworks/*Web*services/library/ws-restful/
[58] http://www.json.org/
[59] Ficheiro com valores separados por vírgulas

solução apresentada, designada por *semantic search service*. O *semantic search service,* disponibiliza duas formas de interação para com o cliente. Como apresentado na Figura 8, é possível verificar a existência de um cliente indicado como aplicação (ex: H-Know) e outro identificado como componente representando um *interface Web* acessível através de um *browser.* O componente *semantic search service* interage com o SPARQL e*ndpoint* do componente virtuoso *universal server,* através de um protocolo http. Por outro lado, o virtuoso, através do componente interno Sponger, efetua uma gestão de *cache* de pesquisas, bem como efetua a agregação e mapeamento na *triple store* local, dados estruturados de um domínio específico, que estejam dispersos pela Internet. O mecanismo de *cache* apresentado e aliado à comunicação direta através de SPARQL *endpoints* a *triple stores* remotas, permite-nos classificar o *semantic search service* segundo a classificação apresentada por [19], e descrita no subcapítulo 3.2, onde são distinguidos três tipos de recuperação de dados. Para este caso em particular, é possível afirmar que a arquitetura se enquadra com o tipo um e dois da classificação.

- Tipo 1: Efetua pesquisa direta em SPARQL *endpoints* sem recurso a qualquer mecanismo de *cache*.
- Tipo 2: Virtuoso *triple store* utilizada no armazenamento de grafos e o Virtuoso Sponger utilizado como *proxy*, com metadados existentes, por exemplo, em páginas *Web*.

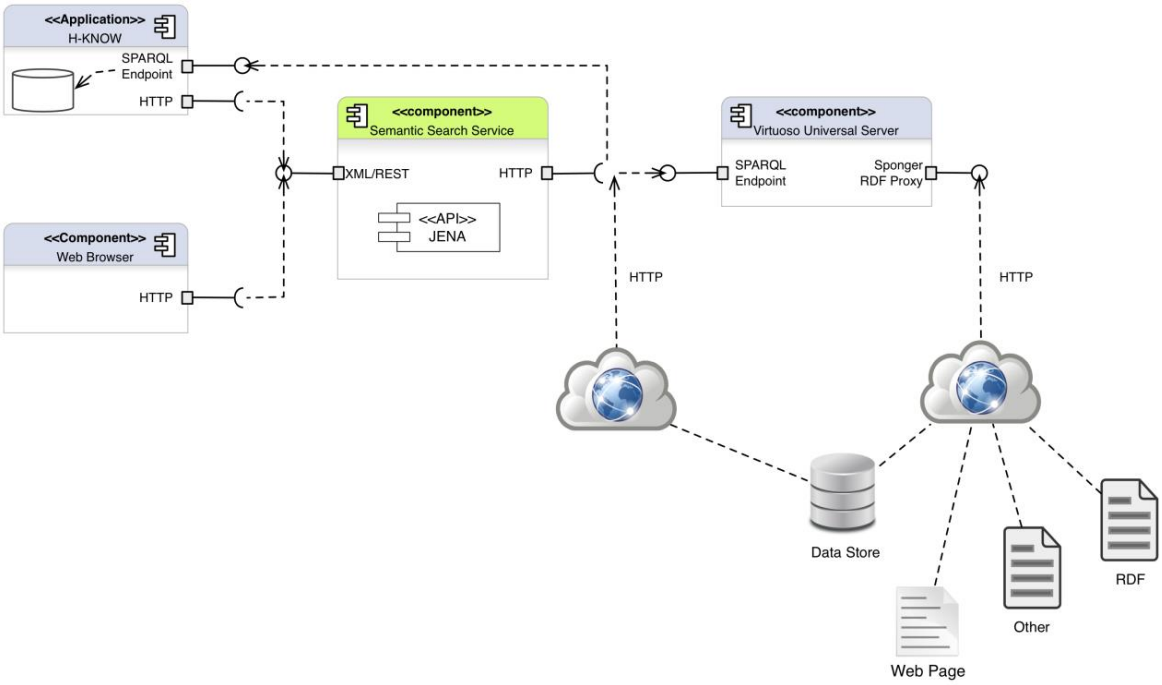


Figura 8:Arquitetura da solução

Internamente ao *semantic search service,* foram implementadas *queries* pré-definidas de pesquisas, onde o Jena é utilizado como *framework* responsável de construir a mensagem a enviar no protocolo de comunicação entre o serviço e a fontes de dados.

Neste tipo de arquitetura, o cliente além de efetuar os pedidos e processar as respostas, no domínio dos dados estruturados, deverá ter especial preocupação nas suas estruturas de representação de dados, respeitando se possível os princípios do *Linked Data,* bem como uma classificação de domínio que seja devidamente utilizada na classificação de conteúdos próprios.

A título de exemplo a Figura 9 apresenta parte do interface que permite a um utilizador especificar (construir) uma pesquisa. A pesquisa pode ser livre, através do campo disponibilizado para o efeito e/ou com base na árvore de conceitos que corresponde à ontologia de domínio da reabilitação de edifícios tendo em conta a herança cultural, que ajuda a compreender e a testar os cenários deste trabalho apresentados no capítulo seguinte.

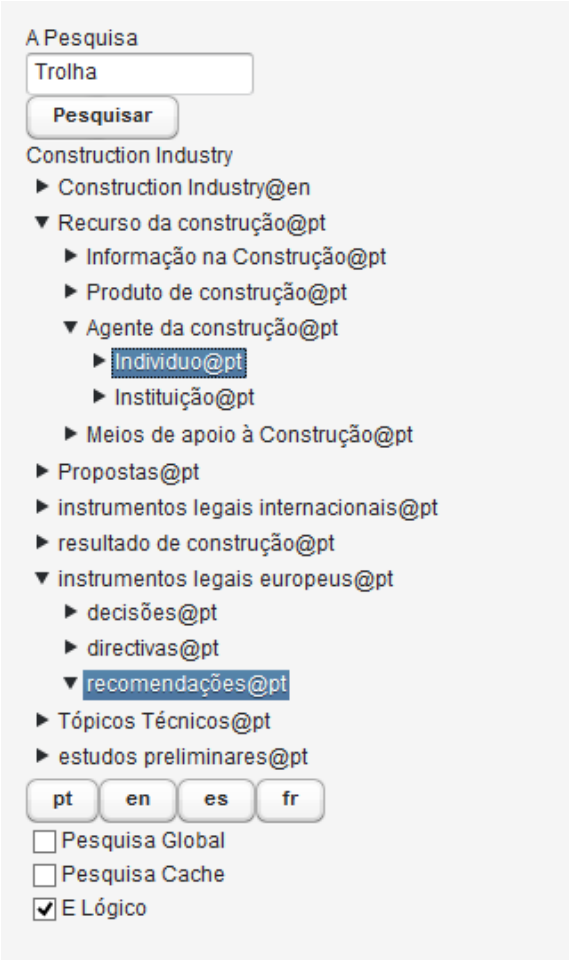


Figura 9:Pesquisa

Considerando a configuração da pesquisa evidenciada pela Figura 9, esta corresponderia a uma pesquisa na *triple store* local, em língua Portuguesa, pelos conceitos "Trolha", "Individuo" e "recomendações", relacionados por um "E" lógico.

4.4.1 Configuração do Serviço

O serviço de pesquisa foi desenvolvido em Java, implementa um conjunto de consultas através da *proxy* e *cache* disponibilizada pelo Virtuoso Sponger, conforme arquitetura apresentada na Figura 8. A cache do *semantic search server,* além de sofrer atualizações mediante as pesquisas efetuadas pelos utilizadores, pode também ter conteúdos adicionados diretamente pelo administrador da plataforma, e encontra-se exemplificado na Figura 10. O padrão típico de um URL, nestas circunstâncias, será:

- **http:// (…)/[fonte_rdf], onde fonte_rdf** , neste exemplo, é http://fr.dbpedia.org.page/Eiffel_(enterprise).

Deste modo temos a possibilidade de adicionar os metadados da página correspondente à **fonte_rdf** às pesquisas em cache do servidor local que suporta o serviço.

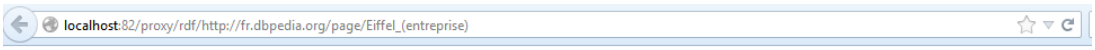


Figura 10:Adição de Dados à cache

Como resultados da correta inserção de dados em *cache*, o administrador deve visualizar no *browser* a seguinte mensagem: "*This document contains 354 facts in HTML Microdata format*" . Ou seja, retorna o número de *tags* que identificam o URI adicionado.

A utilização do virtuoso em processos de configuração por parte do utilizador administrador, não passa só pela adição de dados à *cache*. Para isso, o virtuoso disponibiliza uma interface *Web* de interação com o utilizador, denominado por "*conductor"*, Figura 11, onde o administrador pode aceder a todas as funcionalidades do virtuoso, tais como:

- Administração: atribuição de permissões, *backups*, monitorização e agendamento de processos;
- Gestão da base de dados interna;
- Gestão do servidor de dados (serviços);
- Processador de consultas;
- *Linked Data*: gestão de serviços *Linked Data*, grafos, sponger;
- NNTP (Network News Transfer Protocol) : Ligação com servidores de notícias que respeitam o protocolo NNTP.

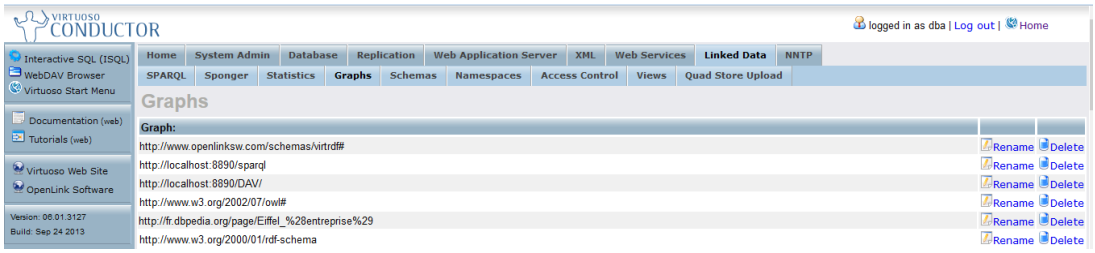


Figura 11:Interface Virtuoso Conductor (Grafos do Virtuoso)

Ao administrador cabe também a tarefa de controlar quantos e quais os *datasets* (SPARQL *endpoints),* com que o serviço pode comunicar. Para isso está disponível um ficheiro *config.xml*, com a estrutura apresentada na Tabela 5. O ficheiro é composto por duas *tags*. A *tag config*, é a *root* do ficheiro XML, e possui uma designação elucidativa para o administrador. A *tag resource* é composta pelo atributo *id*, que existe com o principal objetivo de identificar o recurso identificado através do URI do respetivo SPARQL *endpoint.* Assim, o administrador do serviço tem a possibilidade de limitar as pesquisas dos utilizadores do *semantic search server* a um conjunto de *datasets*, que podem ser definidos como essenciais e de um domínio bem definido.

```
<Config>
<Resource id=”dbpedia”>
http://dbpedia.org/sparql
</Resource>
……
</Config>
```

Tabela 6:Exemplo xml de configuração

Uma vez apresentada a perspetiva de utilização do administrador desta arquitetura, existe a necessidade de direcionar o foco para o lado do cliente. Tal como apresentado na Figura 8, os possíveis clientes do *semantic search service*, poderão ser aplicações colaborativas de dados estruturados, bem como o interface disponibilizado e apresentado na Figura 9. Na perspetiva de um cliente do tipo aplicação, o *semantic search service,* pode ser utilizado para dois tipos de finalidades distintas:

1) Como motor de pesquisa: tal como apresentado anteriormente, ver exemplo da Figura 9, onde o utilizador pode usufruir de um cliente que lhe permita efetuar pesquisas direcionadas;

2) Como serviço de sugestão de conteúdos: uma vez que assumido que as plataformas clientes deste serviço são de dados estruturados. Os resultados das

sugestões são calculados com base nos conceitos que classificam o conteúdo em visualização pelo utilizador.

Assim, para a utilização do *semantic search server*, a aplicação cliente deverá ser capaz de efetuar o seu pedido recorrendo ao protocolo de comunicação RESTFul, onde os pedidos e respostas são transacionados em XML, com a estrutura apresentada na Tabela 6 e Tabela 7 respetivamente.

A Tabela 6, apresenta uma estrutura de uma mensagem de pedido ao *semantic search server,* em XML onde podemos verificar que a sua estrutura é composta por um *root* com quatro atributos utilizados para a aplicação de restrições na pesquisa.

```
<?xml version="1.0" encoding="UTF-8"?>
<pedido lang=$var, local = $loc, cache = $ch,
logica =$log>
<term>
        Construction resource
</term>
<term>
        worker
</term>
</pedido>
```

Tabela 7:XML Input do serviço

Os atributos contidos na *tag* pedido são:

- *Lang*: este atributo identifica a língua na qual o utilizador pretende obter os resultados. Este atributo deve receber as seguintes opções: en (Inglês), pt (Português), es (Espanhol), de (Alemão), fr (Francês) e it (Italiano).
- Atributo Local: Atributo que identifica se a pesquisa deve ser efetuada unicamente à base de dados de suporte ao serviço, ou ao conjunto de *datasets* definidos no ficheiro de configuração (Parâmetro: local/global).
- Atributo *Cache*: Atributo que identifica se a pesquisa irá ser efetuada além da base de dados local, também em ficheiros de *cache* armazenados pelo Sponger (Parâmetro: 1 / 2 , onde 1 corresponde a analisar *cache* e 2 corresponde a não analisar *cache*).
- Atributo *Logica*: Definição da pesquisa baseada num “E” lógico, onde o resultado deve conter referências a todos os termos enviados, ou uma pesquisa baseada num “OU”

lógico, onde os resultados podem retornar referências a um único termo identificado na pesquisa. (Parâmetro: e/ou).

Uma vez efetuadas as restrições ao pedido a efetuar através da parametrização dos atributos da *tag* pedido, existe a necessidade de atribuir valores à *tag term*, que deve ser instanciada, tantas vezes quantos forem os conceitos em pesquisa.

A resposta é contruída pelo *semantic search service,* seguindo a estrutura XML apresentada na tabela 8 e tendo em conta às pesquisas efetuadas, quer pelo mecanismo de *cache* do virtuoso, quer pelos *datasets* definidos pelo administrador

```
<?xml version="1.0" encoding="UTF-8"?>
<Results>
<resource title="Techology Advances" url="http://platform.h-know.eu/?q=forum/techology-advances"> type =”forum”</resource>
</Results>
```

Tabela 8:XML output do serviço

O XML é construído recorrendo a *tags* com nomes intuitivos, de modo a que a interpretação para quem vai efetuar a implementação o processamento de respostas mais rapidamente consiga perceber o conteúdo do XML de resposta. A *tag root* com o *Results*, engloba todos os resultados que deverão ser listados recorrendo á *tag* Resource, composta por três atributos:

- *Title:* O título pode ser interpretado como uma *label* de um determinado conteúdo.
- *Url:* Identifica a localização do conteúdo.
- *type:* Identifica o tipo de conteúdo que está a ser devolvido: fórum, espaço colaborativo, galeria de imagens e página.

Nesta fase do desenvolvimento, começa a ficar claro a importância da usabilidade deste tipo de plataforma. De modo a simplificar a utilização desta solução, foram utilizados *standards* de comunicação, como o RESTFul e XML como “mensageiro”. Esta disponibilização de comunicações e configurações recorrendo a *standards* veem de certa forma ajudar o responsável pela implementação e administração deste serviço, que certamente será uma pessoa com competências informáticas. Na vertente de utilizador cliente do serviço, todos estes processos são transparentes para o utilizador. Este é um simples cliente que usufrui dos serviços disponibilizados.

Durante este capítulo descreve-se o principal artefacto resultante desta dissertação, ou seja, o *semantic search service,* numa perspetiva alicerçada nos pressupostos que sustentam a integração de informação em redes colaborativas, partindo de uma breve apresentação dos requisitos, até à modelação da arquitetura e configurações que lhe estão

subjacentes. O *semantic search service* apresenta-se como um componente inovador desacoplado das arquiteturas informáticas existentes na rede colaborativa e independente do domínio técnico em que essa mesma rede opera. Neste sentido, pode este serviço prevalecer além do ciclo de vida da rede colaborativa e recuperado, posteriormente, numa outra configuração de redes colaborativas de organizações. No capítulo seguinte, apresentam-se diferentes cenários, que para além de procurarem validar a solução aqui descrita, corroboram as caraterísticas imediatamente mencionadas.

Capítulo 5

Teste e Validação da solução

Este capítulo apresenta três cenários de teste para a solução apresentada no capítulo anterior. Para a definição dos cenários de teste foi usada a plataforma colaborativa desenvolvida no âmbito do projeto europeu H-Know. Assim, começa-se por fazer uma breve apresentação do projeto e respetiva plataforma. De seguida é apresentada a abordagem usada para teste e validação da solução, descrevendo o processo e métricas utilizados. De acordo com os resultados obtidos, os mesmos são discutidos e são apresentadas as principais conclusões.

5.1 Projeto H-Know: breve descrição

A plataforma H-Know cuja arquitetura se encontra representada na Figura 12, é o resultado de um trabalho de investigação que resultou na integração e personalização de três plataformas distintas Drupal, Ontowiki e Moodle[60]. Estas 3 plataformas são responsáveis pela implementação de funcionalidades em quatro domínios:

1) ***Knowledge management:*** cujo o principal objetivo passa pela gestão e classificação de conteúdos;
2) ***Management of Social Interactions****:* visa a criação de relações potenciando parcerias de negócio estratégicas, onde os espaços colaborativos possuem um papel essencial;
3) ***Technology-Enhanced Learning:*** possui o intuito de disponibilizar aos utilizadores uma boa oferta formativa;
4) ***Ontology Management****:* permite que a ontologia do domínio utilizada na classificação de conteúdos, se mantenha atual e evolua, com as experiencias de especialistas.

[60] https://moodle.org/

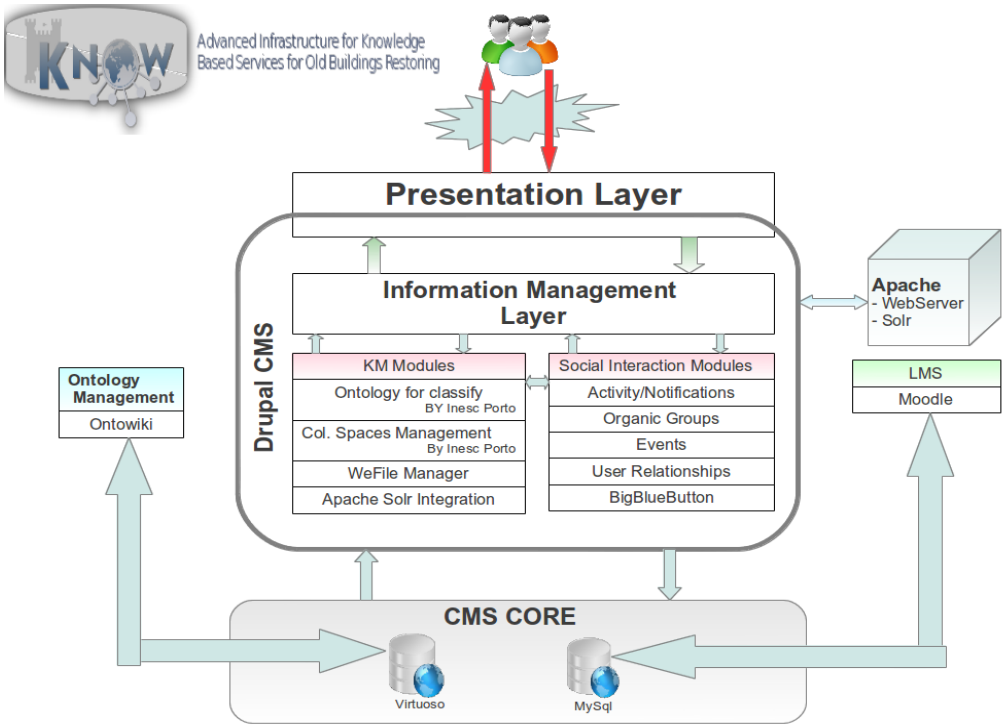


Figura 12:Arquitetura H-Know

Na implementação de funcionalidades de suporte aos dois primeiros domínios apresentados, foi utilizado o Drupal[61]. As funcionalidades associadas ao terceiro e quarto domínios foram desenvolvidas com recurso à plataforma LMS[62] moodle e ontowiki, respetivamente. Nativamente o Drupal é uma plataforma de gestão de conteúdos, considerando os domínios de aplicação do projeto, seria necessário uma plataforma colaborativa de dados estruturados. O desenvolvimento de módulos de suporte à classificação de conteúdos e gestão de espaços colaborativos, foram chave da transformação desta plataforma segundo uma abordagem baseada em *Linked Data.* Existe uma grande comunidade de desenvolvimento, garantindo ao Drupal uma evolução consistente, bem como a manutenção de módulos desenvolvidos. Como podemos analisar na Figura 12, o Drupal é composto por módulos (entenda-se por funcionalidades) nativos orientados à abordagem *Linked Data,* tais como: i) os módulos que permitem a integração com Apache Solr, ii) o mapeamento de informação no vocabulário SIOC e FOAF e RDFs, iii) bem como a classificação de conteúdos utilizando uma ontologia de domínio. Existem também módulos orientados para apoiar a colaboração que permitem a criação e manutenção de espaços

[61] https://drupal.org/
[62] Learning Management System

colaborativos. Nas Figura 13 e Figura 14, é apresentado um espaço colaborativo. Na Figura 13, é possível visualizar que um espaço colaborativo na plataforma H-Know é um espaço virtual, composto por um conjunto de funcionalidades, acessíveis através dos separadores disponibilizados na parte superior do mesmo. São estas funcionalidades que vão ajudar a que o objetivo para o qual o espaço colaborativo foi criado, seja atingido.

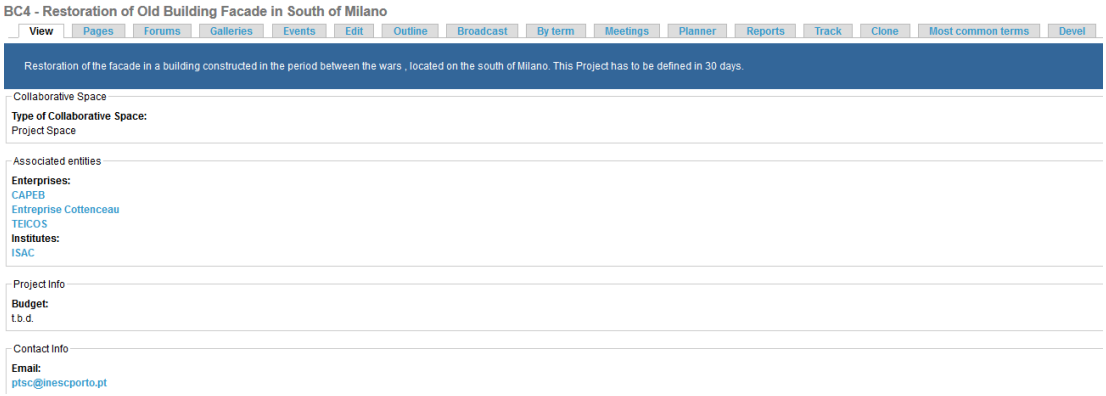


Figura 13:H-Know (Espaços Colaborativos - Ferramentas)

O mecanismo de sugestões utilizado nos espaços colaborativos, podem ser de dois tipos:

1. Tal como apresentado na Figura 14, existe uma coluna do *layout* do espaço colaborativo que apresenta um conjunto de informação, tais como: i) quem está ligado; ii) utilizadores que pertencem ou compõem o espaço colaborativo e; iii) quais os conteúdos adicionados no mesmo. Desta forma o utilizador que se encontre a visualizar o espaço, pode encontrar informação considerada relevante.
2. Um mecanismo de sugestões, que implementa uma pesquisa no repositório de conhecimento da plataforma, por conteúdos que possuam uma classificação semelhante ao espaço/conteúdo em que o utilizador se encontra. O grande desafio proposto para a utilização deste mecanismo, passa pela sensibilização de todos os utilizadores a realizarem a classificação de todos os conteúdos disponibilizados. No entanto, para este trabalho parte-se do pressuposto que os utilizadores procedem à respetiva classificação de conteúdos.

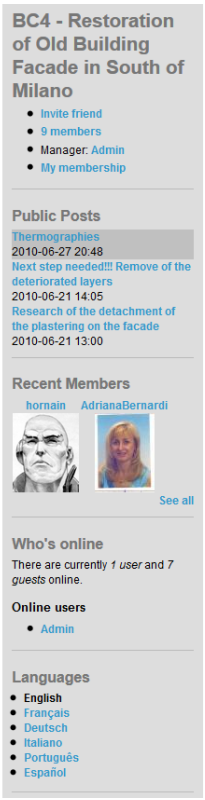


Figura 14: H-Know (Espaços Colaborativos - Gestão)

Focando-nos agora na prespetiva de administração, é apresentada na Figura 15 uma interface de utilização da plataforma ontowiki.

Figura 15:Ontowiki

A ontowiki, foi uma das plataformas abordadas no capítulo três, nomeadamente, identificada como sendo aquela que mais se identificava como proposta para solução do nosso problema. Nesse capítulo foram detetadas fragilidades ao nível de segurança que

levaram à não adoção da mesma plataforma. O facto de a plataforma ser baseada em wiki, a gestão de permissões por conteúdos possui uma abordagem diferente à tradicional, onde não existe o conceito de utilizador com um papel associado. Também é uma plataforma que não privilegia a manipulação de conteúdos como, media e imagens, documentos de texto, entre outros. Contudo, no contexto de utilização de administradores responsáveis pelo desenvolvimento e manutenção da ontologia, onde a comunicação com a *triple store* é determinante, a ontowiki revelou-se uma ferramenta ágil que permite a criação de discussões relativamente à evolução da ontologia e também umo mecanismo de *tracking* que regista todas as alterações efetuadas, permitindo assim que a informação da ontologia nunca desapareça, mas sim seja omitida.

A componente de *technology enhanced learning,* não menos importante no contexto do projeto, não é aqui apresentada, pelo facto de não ser relevante no contexto deste trabalho em particular. Assim, um vez apresentadas as interfaces com utilizador e administrador, temos como suporte a esta plataforma duas bases de dados:

1. Uma base de dados MySQL, partilhada pelo moodle e Drupal. Esta é a responsável pelo armazenamento de todos os conteúdos, perfis de utilizadores e regras/permissões de ambas as plataformas.
2. Uma base de dados RDF (Virtuoso): *Triple store* responsável pelo armazenamento e gestão de metadados da plataforma.

As escolhas apresentadas, trouxeram uma grande flexibilidade em termos de manutenção da plataforma do projeto, bem como para os desenvolvimentos futuros, nos quais se incluí o trabalho desenvolvido nesta dissertação. Esta vantagem deve-se ao facto de, por um lado, a base de dados MySQL ser um produto *opensource,* detido pela ORACLE, que privilegia protocolos de comunicação aceites como *standards*, o que lhe confere grande fiabilidade e usabilidade no suporte ao desenvolvimento em diversos produtos. Por outro lado, o virtuoso é um servidor desenvolvido com base na abordagem *Linked Data*, usado e recomendado pela comunidade científica em [42].

Uma vez apresentada a plataforma colaborativa de dados estruturados que será utilizada na validação da nossa abordagem, foram desenhados três cenários para teste e validação da solução. O cenário um, apresenta o estado atual da plataforma H-Know e foi especificado, no sentido de efetuar uma comparação do antes e depois da implementação do *semantic search service,* resultado desta dissertação.

5.2 Procedimento de teste e validação

Na literatura, são mencionadas diversas abordagens científicas para suportar a avaliação deste tipo de trabalhos. Abordagens de carácter genérico e abordagens de laboratório/controladas [44]. No contexto deste trabalho, as abordagens controladas de avaliação de sistemas IIR são aquelas que merecem principal destaque. Em [44] são apresentadas seis abordagens de avaliação:

- **Abordagem exploratória, descritiva e baseada em estudos:** Este tipo de abordagem tem como principal objetivo a apresentação de uma descrição. É utilizada quando o domínio é pouco conhecido e as questões de investigação são abertas e amplas permitindo também deixar respostas e hipóteses em aberto;
- **Abordagem de experimentação:** Esta abordagem privilegia, por exemplo, a avaliação entre duas plataformas, onde um cenário de teste é aplicado, e o comportamento de ambas é analisado. Esta análise não implica a necessidade de ser um processo executado em simultâneo em duas ou mais plataformas.
- **Abordagem de estudo em laboratório e em ambiente natural:** é uma abordagem que privilegia a avaliação de usabilidade. Os testes efetuados em laboratórios, são realizados quando existe a necessidade de ter controlo sobre variáveis externas que possam ter influência do resultado da avaliação que não é pretendido. Os testes em ambiente natural, contabiliza todas as variáveis que possam ser consideradas "desestabilizadoras" ou não.
- **Abordagem a estudos transversais:** são realizados num período bastante estendido no tempo, onde os critérios de avaliação podem ser definidos por intervalos de tempo.
- **Abordagem aos casos de estudo:** é uma abordagem caraterizada por efetuar um estudo intensivo sobre um determinado caso. Habitualmente são realizados no ambiente natural e utilizadam métodos de abordagens de avaliação transversais. Permite ao analista orientar a sua análise para fatores de avaliação mais determinantes.
- **Abordagens orientadas a simulações:** Este tipo de abordagem, caracteriza-se pelo facto de ser uma experiência controlada, onde o utilizador pode definir os *inputs* e à partida já conhece os *outputs.* Este tipo de abordagem carateriza-se pela atribuição de realismo, quer através da simulação de comportamentos de utilizadores, comunicações e recursos.

Conhecidas as abordagens de experimentação, estando este trabalho inserido no contexto tão exigente como o empresarial, surge a necessidade de efetuar uma validação

prévia da arquitetura. Para isso foram selecionadas abordagens orientadas a ambientes controlados como: simulações e exploratórias.

As simulações permitiram criar ambientes virtuais semelhantes/ou não, ao empresarial, que puderam utilizar *input* definidos e os *outputs,* não colocam em causa as operações da empresa. Neste caso, torna possível a definição de diversos cenários como: cenários ideais ou com variáveis externas de "perturbação" existentes no mundo real. As abordagens exploratórias, permitiram-nos efetuar uma comparação **"as-is"** e **"to-be"**. Isto porque, a existência da plataforma H-Know no cenário de teste, permitiu a criação de comparações entre a plataforma antes do *semantic search service* e a plataforma pós *semantic search service*.

Procedeu-se ainda ao estudo de métricas que nos permitissem avaliar os resultados alcançados. Da revisão da literatura percebeu-se que à dispersão do tipo de métricas de avaliação existentes levou alguns autores, nomeadamente [45], efetuasse uma categorização por: relevância, eficiência, utilidade, satisfação de utilizador e sucesso. Ao longo do tempo estas medidas foram evoluindo, adaptando-se ao aparecimento de motores de busca e em [44], é apresentada uma classificação constituída por quatro classes:

- **Contextual**: caracteriza-se pela utilização de questionários como método de recolha de informação. São efetuadas perguntas de caráter genérico, qual a familiaridade com o sistema em análise? Idade? entre outras. Serve para caracterizar o ambiente onde é efetuada a pesquisa.
- **Interação:** carateriza-se pela análise da interação do utilizador com o sistema, avaliando o número de pesquisas, número de documentos visualizados: Normalmente são dados obtidos através de ficheiros de *log.*
- **Performance:** Caracteriza-se pela análise de *log* da plataforma, onde são registados os conteúdos considerados relevantes. Uma das formas apresentadas para a identificação de conteúdos relevantes, passa pela quantidade de ficheiros que são guardados por pesquisa;
- **Usabilidade:** a usabilidade carateriza-se pelo *feedback* obtido pelos utilizadores.

No contexto deste trabalho, a abordagem utilizada para avaliação do *semantic search service,* passa por uma validação em ambiente controlado, segundo uma abordagem orientada a simulações. A escolha desta abordagem prende-se essencialmente com o tempo para efetuar interações com utilizadores. De salientar que o prazo expectável deste tipo de

desenvolvimento é de cerca de 6 meses. E em ambiente controlado, são definidas um conjunto de casos de teste, tal como os apresentados na secção seguinte.

5.3 Definição dos cenários de teste

5.3.1 Cenário 1

No cenário 1, apresentado neste subcapítulo e esquematizado na Figura 16, éapresentado um caso real de implementação da plataforma H-Know. A título experimental a entidade coordenadora do projeto H-Know, *Fundación Santa Maria de la Real*[63]*(FSMLR),* instalou a plataforma, tendo em vista a criação de valor acrescentado no seu domínio de influência. Assim, segundo [43], a arquitetura desta plataforma traz benefícios em áreas como:

- Identificação de oportunidades de negócio;
- Identificação de especialistas ou instituições conceituadas na área da herança cultural;
- Facilitar gestão e partilha de documentação e informação;
- Facilitar a aquisição de novas competências.

A utilização da plataforma H-Know, instalada em servidores próprios da FSMLR, disponibiliza um número de acessos limitados a associados e potenciais parceiros. A utilização da base de dados relacional MySQL visa no armazenamento de conteúdos e o virtuoso no armazenamento persistente de metadados. Adicionalmente é disponibilizado ao utilizado um mecanismo de pesquisa semântica executada através de *queries* SPARQL à base de dados Virtuoso.

[63] http://www.santamarialareal.org/

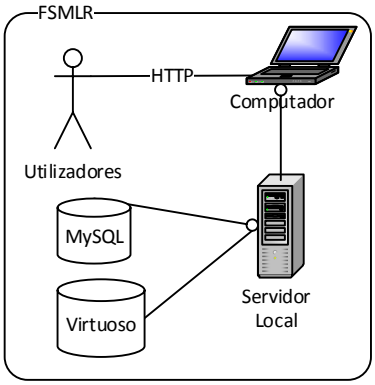


Figura 16: Cenário 1

As pesquisas efetuadas na plataforma H-Know, recorrem a um conjunto de *queries* em SPARQL, definidas e construídas para efetuar consultas na estrutura de dados utilizada bem como, nos critérios de classificação definidos na ontologia. A tabela 9, apresenta um exemplo de uma *query* implementada na plataforma H-Know, onde a cláusula *"where"*, direciona o pedido ao grafo *"hknow".* A *query*, através dos critérios definidos pela cláusula "*where*", restringe os resultados, a um sub-conjunto de grafos da fonte de dados "*hknow*".

```
SELECT  distinct  ?y ?z
 from <hknow>
        WHERE
        {
         ?x rdf:type owl:Class.
          ?x dc:language ?z.
         ?x dc:source ?y.

         FILTER regex(str(?z), "$var").
         filter (langMatches(lang(?z), $lang)).
        }
```

Tabela 9:Consulta ontologia H-Know

A *query* apresentada, efetua uma pesquisa por todas as classes OWL do grafo H-Know[64], que respeitem as condições definidas pelas variáveis $lang, e $var, que representam a língua de pesquisa e o conceito a pesquisar, respetivamente.

[64] Corresponde à integração do vocabulário do domínio (descrito em RDFS) com os vocabulários SIOC e FOAF;

```
SELECT  distinct  ?y ?z
from <hknow>
WHERE {
        ?x rdf:type owl:Class.
        ?x dc:language ?z.
        ?x dc:source ?y.
        ?x dc:title ?w.
FILTER regex(str(?w), '" + $txtSearch + "').
filter (langMatches(lang(?z), '" + $lang + "'))}
```

Tabela 10:Consulta pesquisa livre

A tabela 10 apresenta uma *query*, onde é efetuada uma pesquisa por recursos que estejam identificados com o título da variável *$txtSearch* e com a língua identificada pela variável *$lang*.

```
SELECT *
from <hknow>
   WHERE {
    ?f rdf:type sioc:Forum .
  ?f sioc:id ?s.
FILTER REGEX(str(?s), "^$var")
```

Tabela 11:Pesquisa de Fóruns

Na tabela 11, é efetuada uma consulta por fóruns, que se encontrem identificados com o valor da variável *$var*.

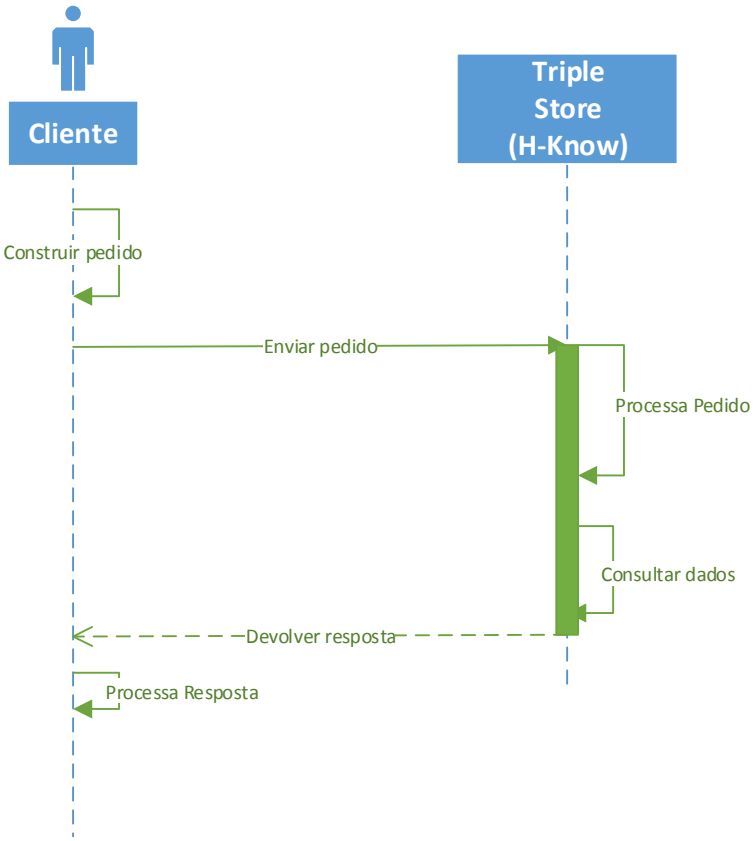


Figura 17:Diagrama de sequência (cenário)

Na sequência das consultas apresentadas, podemos analisar na Figura 17, um diagrama de sequência que apresenta as operações de interação entre objetos. Aqui é possível perceber que a plataforma H-Know se torna uma plataforma "isolada do mundo". A interação de partilha de informação dá-se única e simplesmente entre o utilizador da plataforma e a própria plataforma. Esta plataforma evidenciou-se das demais plataformas, pela suas características inovadoras de gestão de conhecimentos recorrendo a dados estruturados e ontologias, bem como da componente colaborativa assente sobre os espaços colaborativos. Contudo os desenvolvimentos realizados no âmbito deste trabalho potenciados pelos princípios do *Linked Data*, permitem, a sua expansão no contexto de partilha e recuperação de informação da Internet.

5.3.2 Cenário 2

Como apresentado em formato de conclusão do cenário 1, a plataforma H-Know, encontrava-se isolada do mundo. Apesar de ser uma plataforma colaborativa com dados estruturados, a sua ligação com outras plataformas era inexistente. Então é neste cenário que

o *semantic search service* aparece como proposta de serviço que permite a integração de informação entre plataformas colaborativas de dados estruturados.

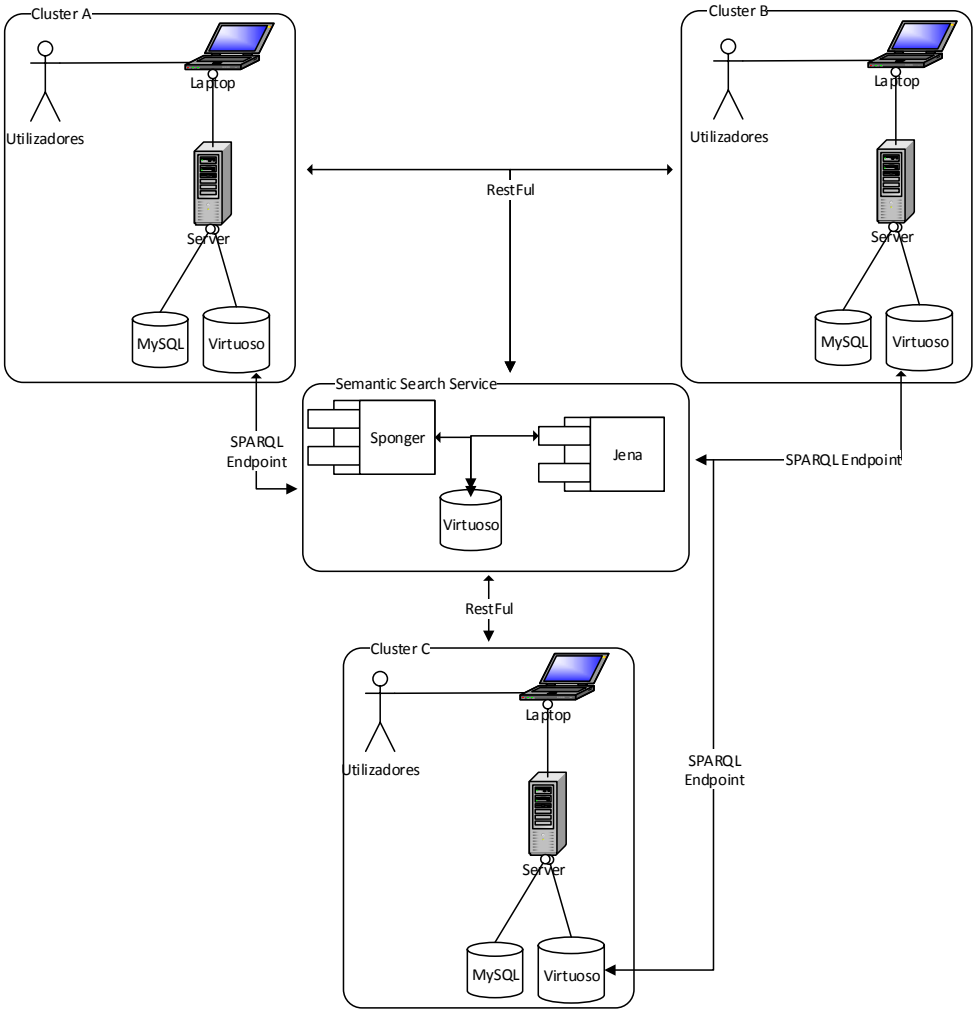


Figura 18:Cenário 2

A implementação deste cenário (Figura 18), implica a existência de duas ou mais plataformas H-Know (Cluster A, Cluster B e Cluster C no exemplo da figura imediatamente anterior). De salientar, que referência à plataforma H-Know existe, pelo facto de ser a plataforma selecionada para a execução dos cenários de teste.

Na elaboração deste cenário, a ontologia do domínio da plataforma H-Know, foi utilizada em diversos clientes, onde cada cliente organizava a sua informação recorrendo à ontologia do domínio da reabilitação de edifícios antigos tendo em conta a herança cultural, e os vocabulários FOAF e SIOC para a identificação das estruturas da organização e pessoas. Este cenário genérico, pode ser particularizado, tendo como ponto de partida o cenário apresentado em [43] e explorando a sua utilização e replicação. Em resumo, o cenário,

considera por exemplo, a *FSMLR,* apresentada no capítulo anterior bem como o *CAPEB*[65], associação Francesa de pequenos empresários da construção civil. Ambas as organizações pretendem efetuar uma implementação da plataforma H-Know em servidores separados, de modo a que cada organização possa garantir aos seus utilizadores uma manutenção e desenvolvimento à medida, bem como a confidencialidade de conteúdos partilhados.

No domínio empresarial, as preocupações de confidencialidade de dados/informação e desenvolvimento à medida são uma constante. Então os conteúdos continuam armazenados nas bases de dados da organização, disponibilizando apenas aqueles que se encontram marcados como conteúdos públicos. Na Figura 19, o diagrama de sequência apresenta o seguimento de operações a realizar para a recuperação de informação de múltiplas *triple stores.* A construção do pedido efetuado pelo cliente, implica a necessidade do desenvolvimento de um cliente do *semantic search service,* capaz de contruir uma mensagem de pedido a enviar através do protocolo RestFul, bem como efetuar a leitura e disponibilização dos conteúdos devolvidos.

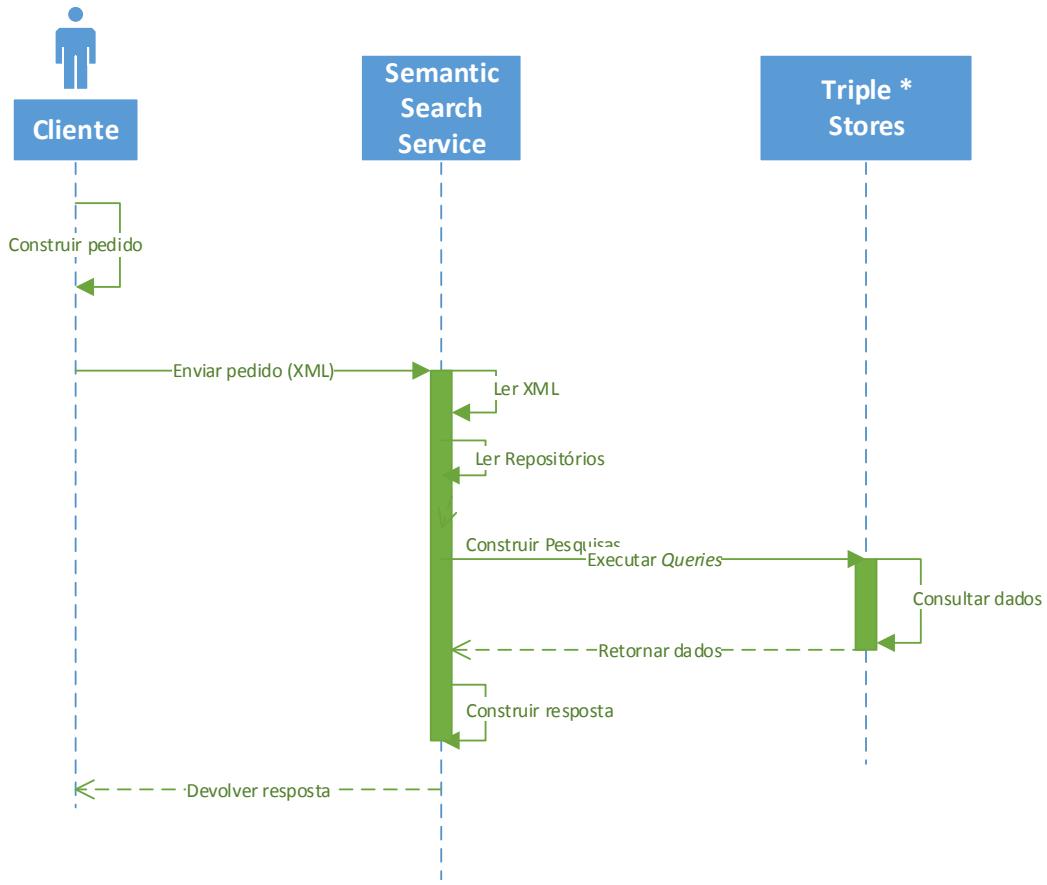


Figura 19:Diagrama de sequência (cenário 2)

[65] http://www.capeb.fr/

Mediante esta pequena contextualização, a aplicabilidade do *semantic search service,* começa a revelar alguns dos seus pontos fortes, que passa pela simplicidade de integração de novas fontes de informação através da adição dessa mesma fonte no ficheiro de configuração apresentado na tabela 6.

```
SELECT  distinct  ?y ?z
        WHERE
        {
         ?x rdf:type owl:Class.
          ?x dc:language ?z.
         ?x dc:source ?y.

         FILTER regex(str(?z), “$var”).
         filter (langMatches(lang(?z), $lang)).
        }
```

Tabela 12: Pesquisa a ontologia do domínio

As *queries* implementadas no *semantic search service*, dispensam a utilização da cláusula *“where”* (ver tabela 12). A ausência da cláusula “*from”* torna-se possível, porque o administrador é responsável pela definição dos SPARQL *endpoints* disponíveis, no ficheiro XML de configuração (ver tabela 6).

O *semantic search service,* apresentado e caraterizado no capítulo 4 como arquitetura de integração de informação entre plataformas colaborativas empresariais, disponibiliza um conjunto de *queries* que permitem um conjunto de pesquisas à componente social da plataforma. Ou seja, são disponibilizadas aos seus utilizadores, mecanismos de pesquisas que permitem efetuar pesquisas por *Blogs*, fórums, pessoas, contatos etc. Para isso o serviço de pesquisa implementa um conjunto de consultas em SPARQL apresentas nas tabelas 13 e 14, direcionadas para a componente colaborativa.

```
SELECT *
   WHERE {
   ?f rdf:type sioc:Forum .
  ?f sioc:id ?s.
FILTER REGEX(str(?s), "^$var")
```

Tabela 13:Pesquisa por forums em SIOC

Na Tabela 13 é possível pesquisar por fóruns que se encontrem identificados com o valor da variável *$var.*

```
SELECT ?mbox
  WHERE
    { ?x foaf:name "Jean Francois" .
      ?x foaf:mbox ?mbox }
```

Tabela 14:Pesquisa FOAF (email)

A tabela 14 apresenta uma consulta por dados de um determinado utilizador (Jean-Francois). Através da identificação *mbox* associada ao prefixo *foaf,* podemos assumir que a pesquisa terá como resultado o *email* desse utilizador.

5.3.3 Cenário 3

O objetivo da apresentação de um terceiro cenário é introduzir um novo componente utilizado no *semantic search service,* o virtuoso Sponger. Este serviço é utilizado para a gestão de um mecanismo de *cache* e recuperação de informação de diversas fontes de dados não *triple stores,* mas sim formatos RDF e Não-RDF.

Assim, efetuando uma breve comparação entre os cenários, dois e três, concluímos que muitos dos tipos de pesquisas, são semelhantes. Assim, a consulta apresentada na tabela 10 deixa de ser exclusiva no contexto da pesquisa livre passando a implementar também a *query* da tabela 15. Na tabela 15, a *query,* representa uma pesquisa aos conteúdos em que a *label* seja composta pelo conteúdo da variável "$var", filtrando os resultados pela língua definida na variável "*lang".*

```
select distinct *

        where {  ?p rdfs:label ?type.
                    FILTER (regex(?type, '$var')).
                FILTER langMatches(lang(?type),'"+lang+"')};
```

Tabela 15: Pesquisa Livre

A representação deste cenário 3 apresentada na Figura 20, complementa a exemplificação apresentada, na medida em que nos permite efetuar uma comparação visual com o cenário dois.

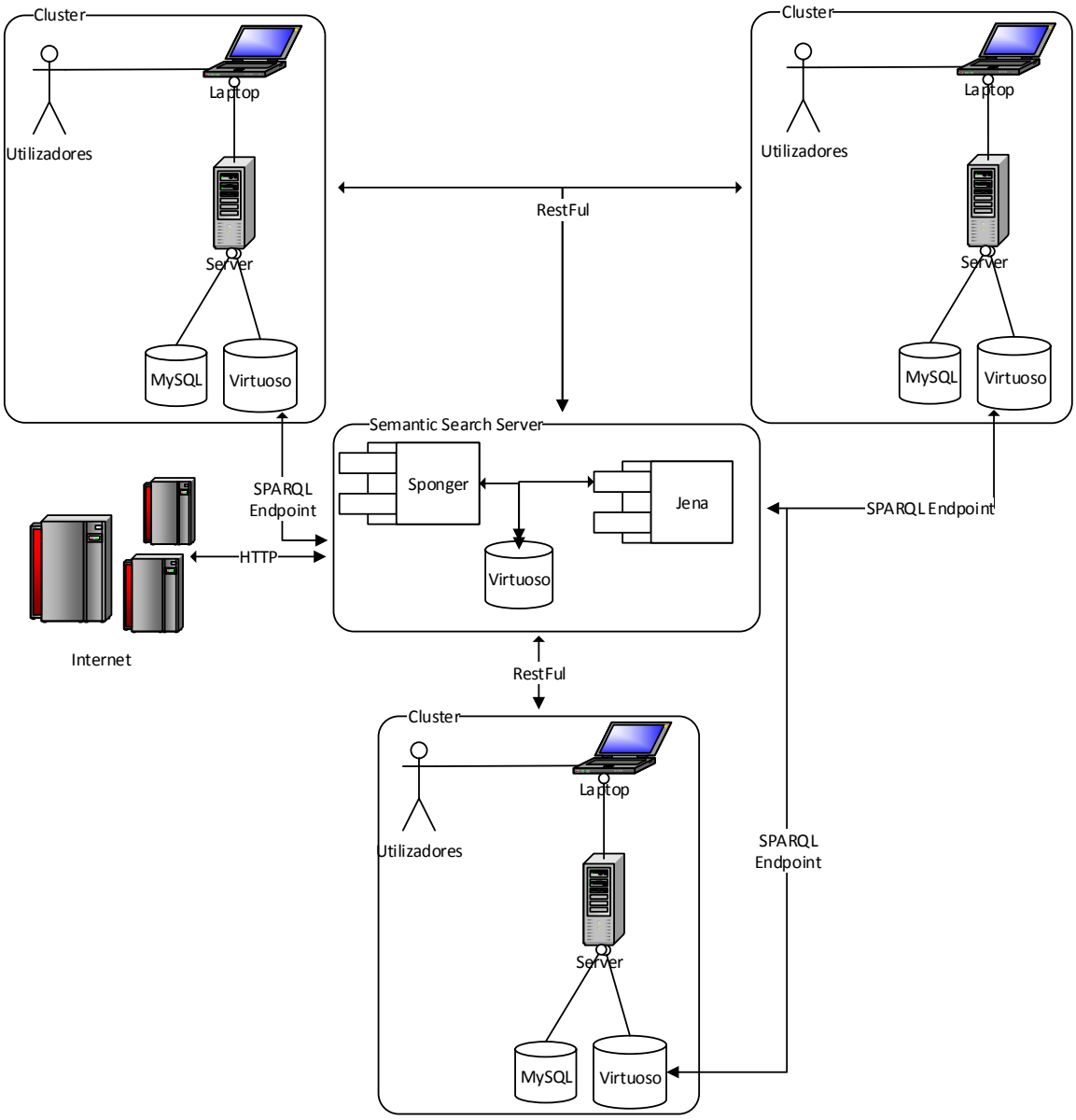


Figura 20: Cenário 3

O diagrama de sequência apresentado na Figura 21, vem ajudar a complementar a interpretação do cenário três apresentado na Figura 20: Existe um conjunto de operações comuns aos diagramas de sequência do cenário anterior, visto que, implementa todas as consultas definidas no cenário anterior acrescentado o componente virtuoso Sponger, que permite a recuperação de dados estruturados RDF e não-RDF.

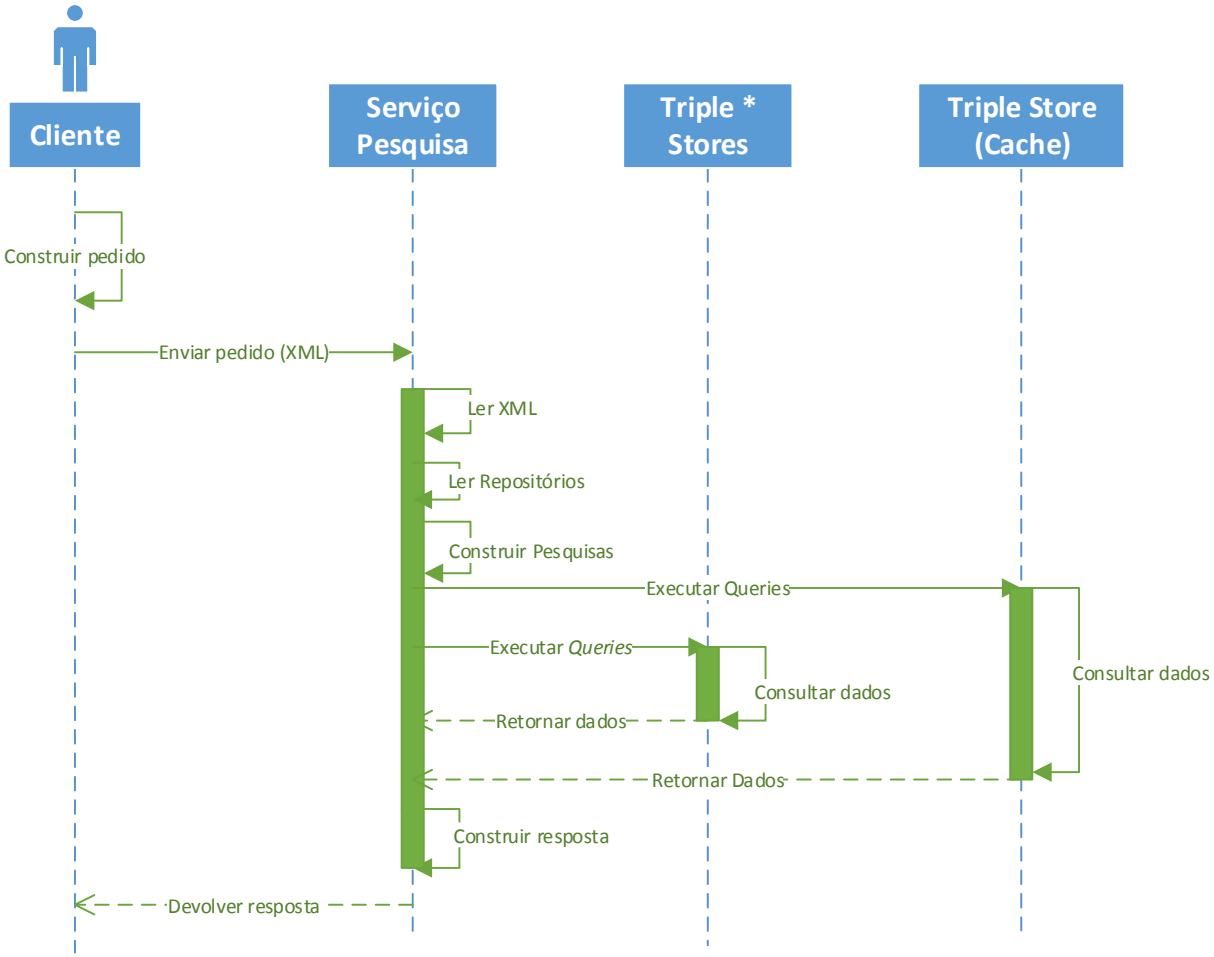


Figura 21: Diagrama de Sequência (Cenário 3)

A utilização deste cenário num contexto real, empresarial, pode ser uma mais valia, na medida em que permite a interação com um maior número de fontes de dados. Contudo, o administrador deverá ter um maior controlo sobre as fontes de dados que vai definir no virtuoso Sponger de modo a não estender em demasiado o domínio das pesquisas.

De notar que neste cenário em particular, o *semantic search service* lida com uma configuração mais heterogénea que os dois primeiros cenários apresentados. É no contexto deste cenário que emerge a versatilidade da arquitetura da solução proposta, em benefício da rede colaborativa. Ou seja, em qualquer altura, e por razões de garantia da operabilidade da rede, se poderá incluir outras organizações com particular interesse para a rede e que, sem disporem de solução própria para gestão e partilha de informação, poderão facilmente integrar a plataforma dispondo de configurações mínimas.

5.4 Resultados

A apresentação de resultados está centrada nos três cenários descritos anteriormente.

Ambas as pesquisas são efetuadas recorrendo à ontologia do domínio da reabilitação de edifícios, desenvolvida no âmbito do projeto H-Know.

Caso de teste 1 - Parâmetros:

- Pesquisa livre: *church*
- Termos da ontologia: *Conservator ; Historian*
- Língua: Inglesa

Caso de teste 2 - Parâmetros:

- Pesquisa: *Structural Reinforcement*
- Termos da ontologia: *Political Information; Directives; Consultant*
- Língua: Inglesa

A caracterização das consultas, identificando o "E" lógico ou não, será definida em cada um dos cenários.

5.4.1 Cenário 1

Aplicando ao cenário os casos de teste apresentados anteriormente obtemos o seguinte mapeamento:

- Em ambas as consultas podemos verificar que a língua é um fator importante, assim na consulta 1 e 2 é identificada a língua da pesquisa que vai atribuir valor à variável *$lang*;
- O texto identificado na consulta como texto livre, atribui valor à variável *$ txtSearch* na consulta apresentada na tabela 10:
- A lista de conceitos da ontologia, atribui valor à variável *$var*, utilizada na consulta da tabela 9, contudo neste caso a utilização da consulta será de forma "recursiva", uma vez que irá ser realizada tantas vezes quanto o número dos conceitos como podemos analisar na tabela 15.

```
For (int i=0; i< concepts.size(); i ++){
        SELECT  distinct  ?y ?z
         from <hknow>
                WHERE
                {
                 ?x rdf:type owl:Class.
                  ?x dc:language ?z.
                 ?x dc:source ?y.

                 FILTER regex(str(?z), “concepts[i]”).
                 filter (langMatches(lang(?z), $lang)).
        }
……
…….
}
```

Tabela 15:Consulta recursiva

Os resultados obtidos em cada iteração, serão comparados com resultados previamente obtidos, de modo a não existir redundância de dados.

Caso de teste 1 - Resultados:

Titulo	URI	Tipo
Page 1: Introduction. History of San Roque Church	http://platform.h-know.eu/?q=page/page-1-introduction-history-san-roque-church	Page
Page 2: Architecture of San Roque Church	http://platform.h-know.eu/?q=page/page-2-architecture-san-roque-church	Page
Reinforcement structures	http://platform.h-know.eu/?q=gallery/reinforcement-structures	Gallery
Ventimola	http://platform.h-know.eu/?q=enterprise/ventimola-gmbh-co-daemmtechnik-kg	Enterprise
……..	……	

Tabela 16: Resultados Cenário 1, Consulta 1

Esta pesquisa efetuada a aplicada na *triple store* do h-know da plataforma H-Know do servidor da FSMLR foram devolvidos 24 resultados, onde 18 foram se encontram identificados como páginas de Internet, 2 são galerias de fotos e 4 empresas.

Caso de Teste 2 - Resultados:

Titulo	URI	Tipo
Doe Structural Reinforcement	http://platform.h-know.eu/?q=project/doe-structural-reinforcement	Collaborative Spaces
Page 2: Architecture of San Roque Church	http://platform.h-know.eu/?q=page/page-3-archaeological-discoveries	Page
Reinforcement structures	http://platform.h-know.eu/?q=gallery/reinforcement-structures	Gallery
Entreprise Cottenceau	http://platform.h-know.eu/?q=enterprise/entreprise-cottenceau	Enterprise
……..	……	

Tabela 17:Resultados Cenário 1, Consulta 2

Na consulta 2, os valores a considerar nesta pesquisa, passam por 18 resultados, onde 12 são páginas *Web*, 2 espaços colaborativos, 1 galeria de imagens e 3 empresas. Como podemos analisar, nos valores de alguns atributos devolvidos na pesquisa, temos resultados que se repetem. Esta situação acontece não só por mera coincidência de classificações idênticas, mas também porque a lógica da pesquisa implementada no projeto recorre a um, "ou" lógico. Uma vez que a plataforma se encontra em fase de lançamento e a quantidade de informação lá colocada, em crescimento, era expectável que um grande número de pesquisas fosse ter resultado nulo.

5.4.2 Cenário 2

Aplicando a este cenário os dois casos de teste definidos anteriormente, obtemos os seguintes resultados:

Caso de teste 1 - Resultados:

Titulo	URI	Tipo
Page 1: Introduction. History of San Roque Church	http://platform.h-know.eu/?q=page/page-1-introduction-history-san-roque-church	Page
Page 2: Architecture of San Roque Church	http://platform.h-know.eu/?q=page/page-2-architecture-san-roque-church	Page
Page 3: Archaeological discoveries	http://portal.h-know.eu/?q=page/page-3-archaeological-discoveries	Page
Meeting1	http://portal.h-know.eu/?q=event/meeting-1	Page
……..	……	

Tabela 18: Resultados Cenário 2.1, Consulta 1

Nesta consulta, podemos analisar que:

- Obtemos todos os resultados da consulta efetuada no cenário anterior, mais os resultados da plataforma utilizada pelo CAPEB.
- Foram obtidos mais resultados: devolvidos 30 resultados, onde 24 foram se encontram identificados como páginas de Internet, 2 são galerias de imagens e 4 empresas

Caso de teste 2 – Resultados:

Titulo	URI	Tipo
Doe Structural Reinforcement	http://platform.h-know.eu/?q=project/doe-structural-reinforcement	Collaborative Spaces
Page 2: Architecture of San Roque Church	http://platform.h-know.eu/?q=page/page-3-archaeological-discoveries	Page
Page 4: Intervention Stage	http://portal.h-know.eu/?q=page/page-4-intervention-stage	Gallery
Entreprise Cottenceau	http://platform.h-know.eu/?q=enterprise/entreprise-cottenceau	Enterprise
……..	……	

Tabela 19:Resultados cenário 2, Consulta2

Neste cenário esta consulta, em termos quantitativos foi muito pouco vantajosa, na medida em que, só aumentou face ao cenário anterior, mais um resultado. Assim, obtivemos 19 resultados, onde 13 são páginas *Web*, 2 espaços colaborativos, 1 galeria de imagens e 3 empresas.

Como podemos analisar, nos URL dos resultados obtidos, ambas as plataformas encontram-se a partilhar o mesmo domínio, visto que, como já foi identificado no início deste capítulo, estas são plataformas que se encontram em fase de testes, sendo que alguns das páginas, espaços e ou fóruns se encontram disponíveis com informação meramente identificativa.

5.4.3 Cenário 3

Este cenário implementa todas as consultas apresentadas nos cenários anteriores, assim é expectável que apresente os resultados encontrados em pesquisas anteriores mais s resultados as pesquisas efetuadas recorrendo à DBpedia (neste caso).

Caso de teste 1 - Resultados:

Titulo	URI	Tipo
Page 1: Introduction. History of San Roque Church	http://platform.h-know.eu/?q=page/page-1-introduction-history-san-roque-church	Page
Luke the Historian	http://dbpedia.org/resource/Luke_the_Historian	Page
Page 3: Archaeological discoveries	http://portal.h-know.eu/?q=page/page-4-intervention-stage	Page
Roman Catholic churches	http://dbpedia.org/resource/Category:Roman_Catholic_churches	Page
……..	……	

Tabela 20:Resultados cenário 3, consulta1

Os resultados obtidos nesta consulta possuem uma grande expressão, uma vez que foram devolvidos 4241 resultados.

Caso de teste 2 - Resultados:

Titulo	URI	Tipo
Doe Structural Reinforcement	http://platform.h-know.eu/?q=project/doe-structural-reinforcement	Collaborative Space
Cambridge Consultants	http://dbpedia.org/resource/Cambridge_Consultants_Ltd	Page
Page 4: Intervention Stage	http://portal.h-know.eu/?q=page/page-3-archaeological-discoveries	Page
Environmental Consultant	http://dbpedia.org/resource/Environmental_Consultant	Page
……..	……	

Tabela 21: Resultados cenário 3, consulta2

O número de resultados desta consulta é menos expressivo, uma vez que estamos a considerar uma pesquisa por expressões compostas normalmente por dois conceitos, restringindo obrigatoriamente a consulta pelos termos nos repositórios de dados. O número total de resultados obtidos foram 155 resultados, onde 149 são páginas *Web*, 2 espaços colaborativos, 1 galeria de imagens e 3 são empresas.

5.4.4 Principais conclusões

Como conclusão deste capítulo são apresentados os resultados da utilização do serviço de pesquisa recorrendo *Linked Data* quando aplicado a dois casos de testes diferentes. A avaliação segundo a performance no caso teste 1, é apresentada na tabela 22, onde se pode analisar um comparativo de resultados obtidos entre os três cenários. De destacar, que no cenário três a quantidade de resultados é substancialmente superior face aos restantes.

Tipo de resultados	Cenário 1	Cenário 2	Cenário 3
Página *Web*	18	24	4235
Espaços colaborativos	2	2	2
Galeria de Imagens	x	x	x
Empresas	4	4	4

Tabela 22:Avaliação da Performance (Caso teste 1)

Na tabela 23, apresenta o comparativo entre cenários, segundo o critério de pesquisa definido no caso de teste 2.

Tipo de resultados	Cenário 1	Cenário 2	Cenário 3
Página *Web*	12	13	149
Espaços colaborativos	2	2	2
Galeria de Imagens	1	1	1
Empresas	3	3	3

Tabela 23: Avaliação performance (caso teste 2)

De modo a efetuar uma análise mais detalhada dos comparativos, que visam a análise de performance, é possível fazer uma divisão dos cenários, analisando os resultados

individualmente. Esta análise, foi efetuada recorrendo aos três cenários de teste e aplicando os casos de testes já apresentados.

Ao cenário 1 que representa o estado atual da plataforma, foram aplicados ambos os casos de teste definidos para validação dos cenários. Os resultados obtidos foram os seguintes (tabela 24):

Tipo de resultados	Cenário 1	
	Caso Teste 1	**Caso Teste 2**
Página *Web*	18	12
Espaços colaborativos	2	2
Galeria de Imagens	0	1
Empresas	0	3

Tabela 24:Resultados Cenário 1

Aplicando o mesmos critérios de pesquisa no cenário 2 que, recorre ao *semantic search service* e, permite ao utilizador efetuar uma consulta, que obtém informação da própria plataforma e de plataformas externas. Foram obtidos os resultados apresentados na tabela 25 e 26.

Tipo de resultados	*FSMLR*	
	Caso Teste 1	**Caso Teste 2**
Página *Web*	18	12
Espaços colaborativos	2	2
Galeria de Imagens	0	1
Empresas	4	3

Tabela 25: Resultados plataforma H-Know

Tipo de resultados	*CAPEB*	
	Caso Teste 1	**Caso Teste 2**
Página *Web*	6	1
Espaços colaborativos	0	0
Galeria de Imagens	0	0
Empresas	0	0

Tabela 26 :Resultados plataforma CAPEB

Aplicando os mesmos caso de teste no cenário 3, que é um cenário ainda mais complexo, uma vez que permite a recuperação de informação de fontes externas como DBPedia, FreeBase, entre outros, o número de resultados é muito superior (ver tabela 27).

Tipo de resultados	*DBPedia*	
	Caso Teste 1	**Caso Teste 2**
Página *Web*	4211	136
Espaços colaborativos	0	0
Galeria de Imagens	0	0
Empresas	0	0

Tabela 27: Resultados DBPedia

Analisando os resultados separadamente por plataforma, com os cenários onde as plataformas se encontram integradas recorrendo ao *semantic search service,* por exemplo no cenário 2, onde os utilizadores da plataforma do CAPEB podem aceder a contatos de empresas que são apresentadas com os critérios de pesquisas definidos no casos de teste.

O critério definido pela usabilidade, neste tipo de experiência controlada baseada em simulação, é difícil de avaliar. Neste sentido, de notar que ao nível de administração o *semantic search service,* possui um ficheiro de configuração de fácil compreensão, onde facilmente o administrador consegue atribuir novas fontes de informação a disponibilizar no serviço de pesquisa.

Após a análise dos resultados obtidos, verifica-se que existem mais-valias na utilização da solução apresentada nesta dissertação. Em forma de síntese, estas mais valias são:

- Implementa os princípios *Linked Data;*
- Utilização como ferramenta de pesquisas e/ou sugestões;
- Integração com vocabulários que suportam redes colaborativas (FOAF e SIOC);
- Maior agilidade na integração de plataformas empresariais;
- Facilidade de configuração;

Capítulo 6

Conclusões e Trabalho Futuro

6.1 Conclusões

Como resultado deste trabalho, procurou-se construir uma solução de engenharia, capaz de apoiar a ligação de plataformas colaborativas de dados estruturados, tanto genéricas, em domínios específicos.

Como ponto de partida, foi necessário efetuar um estudo sobre plataformas e abordagens que respondem-se às necessidades de pesquisa de informação em ambientes colaborativos, envolvendo diversas organizações, identificadas na sequência do trabalho desenvolvido no âmbito do projeto europeu H-Know. Deste estudo concluiu-se que nenhuma das plataformas existentes conseguia dar a resposta desejada, sendo necessário desenhar uma nova arquitetura. Percebendo-se que este já era um domínio bastante explorado, onde já existem muitos desenvolvimentos, a principal preocupação focou-se na seleção das ferramentas e tecnologias que poderiam ser utilizadas na resposta a esta necessidade.

Como principais conclusões, interessa salientar os seguintes aspetos: (1) a partilha de dados de forma estruturada deve ser algo em que se deve apostar e (2) deve-se incentivar as organizações à utilização deste tipo de plataformas de dados estruturados, fornecendo-lhes funcionalidades que lhes facilitem o acesso a informação relevante de forma rápida e eficaz.

Ficou também evidente que embora o *Linked Data* seja uma abordagem fundamental na gestão e partilha de informação, os mecanismos que garantem a qualidade de dados são ainda algo rudimentares.

Na nossa opinião, e após a análise dos dados obtidos através dos três cenários de teste deste trabalho, o *Linked Data* deveria crescer aleando-se à comunidade de *Data Mining*,

juntando o que de melhor existe na publicação e partilha de informação, como o melhor do domínio do tratamento da qualidade dos dados.

Estamos certos também, que deveria existir por parte dos utilizadores uma preocupação na classificação dos seus conteúdos nas plataformas de informação que utilizam. Isto porque, como podemos analisar nos cenários apresentados, os resultados poderiam ter sido bem melhores, considerando que ambas as plataformas possuem um número razoável de utilizadores, e no entanto não existe grande cuidado na classificação dos conteúdos.

6.2 Possibilidades de Trabalho Futuro

Este trabalho, serve como ponto de partida para possíveis projetos de investigação que poderão conjugar os conceitos de *Data Mining*, *Linked Data*, bem como análise semântica de conceitos. A análise semântica de conceitos foi já aplicada ao cenário três, nomeadamente, no processamento da informação que resultou de consultas em SPARQL *endpoints* remotos.

Sugere-se também uma exploração do cenário três, tendo em conta a utilização de novas consultas, nomeadamente recorrendo a SPARQL *endpoints*, que utilizem outras ontologias, representadas por modelos distintos dos utilizados neste cenário.

Para a utilização da solução apresentada nesta dissertação principalmente no cenário dois, deve existir um trabalho de sensibilização dos utilizadores das plataformas de dados estruturados na classificação dos conteúdos publicados. Para isso, devem ser utilizados cenários de teste constituídos por especialistas do domínio, que contribuam com informação relevante, demonstrando assim, que a classificação de conteúdos é potenciadora da geração de valor acrescentado nas consultas efetuadas.

Referências Bibliográficas

[1] H.-K. Consortium, “H-Know Objectives,” 2010. .

[2] J. F. F. Dorogovtsev, S. N, Mendes, “Evolu-tion of Networks: From Biological Nets to the Inter-net and WWW,” 2003.

[3] L. M. Cam and A. R. In, “Collaborative networks : a new scientific discipline,” pp. 439–452, 2005.

[4] B. Rubenstein-montano, J. Liebowitz, J. Buchwalter, and D. Mccaw, “A systems thinking framework for knowledge management,” 2001.

[5] M. Alavi, “KNOWLEDGE MANAGEMENT,” vol. 1, no. February, pp. 1–37, 1999.

[6] T. Berners-Lee, “Linked Data Design Issues,” *2009*. .

[7] *NETWORK-CENTRIC COLLABORATION AND SUPPORTING FRAMEWORKS IFIP - The International Federation for Information Processing*. .

[8] “No TitleNational Research Council (1998). Visionary manufacturing challenges for 2020, visionary manufacturing challenges for 2020/committee on visionary manufacturing challenges. In Board on manufacturing and engineering design commission on engineering .” [Online]. Available: http://books.google.pt/books?id=qDYrAAAAYAAJ&printsec=frontcover&hl=pt-PT&source=gbs_ge_summary_r&cad=0#v=onepage&q&f=false.

[9] L. M. Camarinha-Matos, H. Afsarmanesh, N. Galeano, and A. Molina, “Collaborative networked organizations – Concepts and practice in manufacturing enterprises,” *Comput. Ind. Eng.*, vol. 57, no. 1, pp. 46–60, Aug. 2009.

[10] P. Dillenbourg, M. Baker, A. Blaye, and C. O. Malley, “The evolution of research on collaborative learning,” pp. 189–211, 1996.

[11] R. Holliman and E. Scanlon, “Investigating cooperation and collaboration in near synchronous computer mediated conferences,” *Comput. Educ.*, vol. 46, no. 3, pp. 322–335, Apr. 2006.

[12] A. L. Soares and F. Alves, “Collaborative Spaces as Mediators for Information Sharing in Collaborative Networks,” 2012.

[13] L. M. Camarinha-matos and H. Afsarmanesh, “Towards a reference model for collaborative networked organizations,” vol. 06, pp. 4–6.

[14] A. L. Soares and F. J. Alves, “Collaborative Spaces as Mediators for Information Sharing in Collaborative Networks,” 2012.

[15] C. Carneiro, A. L. Soares, and C. Sousa, “Integration of domain and social ontologies in a CMS based collaborative platform,” *OTM*, 2010.

[16] Binyam Chakilu Tilahum, “Linked Data based Health Information Representation , Visualization and Retrieval System on the Semantic Web,” 2013.

[17] C. F. Da Silva, L. Médini, S. A. Ghafour, P. Hoffmann, P. Ghodous, and C. Lima, “Semantic Interoperability of Heterogeneous Semantic Resources,” *Electron. Notes Theor. Comput. Sci.*, vol. 150, no. 2, pp. 71–85, Mar. 2006.

[18] A. P. Sheth, “IN INFORMATION SYSTEMS : FROM SYSTEM , SYNTAX , STRUCTURE TO SEMANTICS FOCUS ON,” 1998.

[19] L. G. Nardin, A. a. F. Brandão, and J. S. Sichman, “Experiments on semantic interoperability of agent reputation models using the SOARI architecture,” *Eng. Appl. Artif. Intell.*, vol. 24, no. 8, pp. 1461–1471, Dec. 2011.

[20] P. Haase, J. Broekstra, A. Eberhart, and R. Volz, “A Comparison of RDF Query Languages,” 2001.

[21] P. Hayes, “RDF Semantics, W3C Recommendations,” 2004. [Online]. Available: http://www.w3.org/TR/2004/REC-rdf-mt-20040210/#rules.

[22] B. C. Tilahun, “Linked Data based Health Information Representation , Visualization and Retrieval System on the Semantic Web,” 2013.

[23] P. J. Groen and M. Wine, “Medical Semantics, Ontologies, Open Solutions and EHR Systems,” 2009. .

[24] T. Eiter, G. Ianni, A. Polleres, R. Schindlauer, H. Tompits, U. Rey, and J. Carlos, “Reasoning with Rules and Ontologies,” 2003.

[25] S. Bechhofer, F. van Harmelen, J. Hendler, I. Horrocks, D. L. McGuinness, P. F. Patel-Schneider, and L. Andrea Stein, “OWL Web Ontology Language Reference,” 2004. [Online]. Available: http://www.w3.org/TR/owl-ref/.

[26] M. Menárguez-Tortosa and J. T. Fernández-Breis, “OWL-based reasoning methods for validating archetypes.,” *J. Biomed. Inform.*, vol. 46, no. 2, pp. 304–17, Apr. 2013.

[27] Tim Berners-Lee, “Linked Data- Design issues,” 2006. [Online]. Available: HTTP://www.w3.org/DESIGNISSUES/LINKEDDATA.HTML.

[28] P. Champin and U. De Lyon, “User Assistance for Collaborative Knowledge Construction,” 2012.

[29] V. Nebot and R. Berlanga, “Building data warehouses with semantic *Web* data,” *Decis. Support Syst.*, vol. 52, no. 4, pp. 853–868, Mar. 2012.

[30] S. Bechhofer, F. van Harmelen, J. Hendler, I. Horrocks, D. L. McGuinness, P. F. Patel-Schneider, and L. Andrea Stein, “http://www.w3.org/TR/owl-ref/,” 2004. .

[31] P. Frischmuth, J. Klímek, S. Auer, S. Tramp, and J. Unbehauen, “Linked Data in Enterprise Information Integration,” vol. 0, pp. 1–17, 2012.

[32] S. Schaffert, J. Eder, S. Gr, T. Kurz, R. Sint, and S. Stroka, “KiWi – A Platform for Semantic Social Software.”

[33] S. Auer, “Powl – A Web Based Platform for Collaborative Semantic Web Development,” pp. 1–10.

[34] S. Dietzold and T. Riechert, “OntoWiki – A Tool for Social , Semantic Collaboration.”

[35] S. Taylor, N. Jekjantuk, C. Mellish, and J. Z. Pan, “Reasoning Driven Configuration of Linked Data Content Management Systems.”

[36] T. Schandl, A. Blumauer, and N. Gmbh, “PoolParty : SKOS Thesaurus Management utilizing Linked Data.”

[37] S. Schaffert, T. Kurz, C. Bauer, F. Dorschel, and M. Fernandez, “The Linked Media Framework Integrating and Interlinking Enterprise Media Content and Data,” 2012.

[38] J. Kotowski, “A Perfect Match for Reasoning , Explanation , and Reason Maintenance : OWL 2 RL and Semantic Wikis,” pp. 1–5.

[39] J. Fernandes, D. Duarte, C. Ribeiro, C. Farinha, J. M. Pereira, and M. M. Da Silva, “iThink: A Game-Based Approach Towards Improving Collaboration and Participation in Requirement Elicitation,” *Procedia Comput. Sci.*, vol. 15, pp. 66–77, Jan. 2012.

[40] F. Schneider and B. Berenbach, “A Literature Survey on International Standards for Systems Requirements Engineering,” *Procedia Comput. Sci.*, vol. 16, pp. 796–805, 2013.

[41] D. Smiley and E. Pugh, *Solr 1.4 Enterprise Search Server*. 2009.

[42] C. Cuijpers, “Legal aspects of open source intelligence – Results of the VIRTUOSO project,” *Comput. Law Secur. Rev.*, vol. 29, no. 6, pp. 642–653, Dec. 2013.

[43] A. Egusquiza and J. L. Izkara, “H-KNOW: Advanced infrastructure for knowledge based services for buildings restoring,” *2012 18th Int. Conf. Virtual Syst. Multimed.*, pp. 461–467, Sep. 2012.

[44] D. Kelly, “Methods for Evaluating Interactive Information Retrieval Systems with Users,” *Found. Trends® Inf. Retr.*, vol. 3, no. 1—2, pp. 1–224, 2007.

[45] L. T. Su, “Evaluation Measures for Interactive Information Retrieval,” *Inf. Process. Manag.*, vol. 28, no. 4, pp. 503–516, 1992.

[46] A. Hevner and S. Chatterjee, “Design Research in Information Systems,” vol. 22, pp. 109–119, 2010.