

RETRIEVO 3

AGREGAÇÃO DE CONTEÚDOS E PESQUISA FEDERADA

CARACTERÍSTICAS E REQUISITOS
TÉCNICOS

SOBRE O DOCUMENTO

Identificador	WP191402		
Aprovado por	Luís Miguel Ferros	Aprovado em	2020-05-15
Classificação	Público		
Distribuição	N/A		

REVISÕES

#	Data	Autor	Alterações
1	2019-12-01	Márcia Fernandes	Documento inicial
2	2020-05-08	Hélder Silva	Atualização e revisão do documento
3	2020-05-08	Miguel Ferreira	Revisão
4	2020-05-15	Luís Miguel Ferros	Revisão final

SUMÁRIO EXECUTIVO

O Retrievio é um software que permite recuperar informação de diferentes fontes de informação existentes numa organização. Por exemplo, o Retrievio pode ser utilizado para pesquisar simultaneamente no catálogo da biblioteca, no sistema de gestão documental e no sítio Web da instituição através de uma única interface de pesquisa.

Este documento tem como objetivo descrever as principais características, funcionalidades e vantagens do software Retrievio, bem como os requisitos técnicos necessários à sua correta instalação.

**PONTO DE
ACESSO ÚNICO
A TODOS OS
RECURSOS DE
INFORMAÇÃO
DA SUA
ORGANIZAÇÃO**

RETRIEVO

O Retrievio é um software que permite implementar portais de pesquisa e recuperação de informação sobre qualquer fonte de informação existente numa organização. Por exemplo, este poderá ser usado para pesquisar simultaneamente no catálogo da biblioteca, no sistema de gestão documental e no sítio Web da instituição através de uma única interface de pesquisa.

O Retrievio não tem como objetivo substituir as fontes de informação existentes, mas sim fornecer um ponto de acesso único e privilegiado à informação que se encontra distribuída pelos diferentes quadrantes da organização.

Após localizar a informação no Retrievio, o utilizador é redirecionado para o sistema que contém a informação primária, permitindo-lhe consultar a respetiva informação de forma completa e no seu contexto de produção.

Para além de permitir a pesquisa centralizada de informação, este software atua também como fornecedor de informação para outros sistemas, podendo ser integrado com agregadores de informação internacionais como a Europeana, o Portal Europeu de Arquivos, o Portal OpenAire ou outros serviços de pesquisa federada como a B-on.

Informação acessível a partir de um **único ponto de acesso**

O Retrievio permite localizar informação que se encontra dispersa por diferentes sistemas de informação, mesmo quando estes sistemas parecem, à partida, incompatíveis.

A pesquisa é sempre realizada através de uma interface comum, evitando a necessidade do utilizador aceder a cada um dos sistemas individualmente.

Aumento da **visibilidade** da instituição

Além de consolidar num único local a informação de vários sistemas internos, o Retrievio permite disponibilizar a informação consolidada e limpa a serviços externos de recolha de informação.

Estes serviços permitem criar valor acrescentado à informação, como por exemplo, serviços adicionais de localização e pesquisa, criação de portais temáticos, estatísticas, etc.

Uniformização de sistemas heterogéneos

O Retrievio é compatível com vários protocolos de comunicação bem estabelecidos tais como o OAI-PMH, Z39.50 e SRU, bem como APIs REST, SOAP, conectores SQL e um conjunto alargado de gateways de acesso a bases de dados de publicações científicas (e.g. EBSCO, ABI, ProQuest, etc.).

Independentemente do protocolo utilizado ou do sistema que suporta a informação, o Retrievio será sempre capaz de a recolher, limpar, indexar e apresentar convenientemente ao utilizador final.

Ativos de **informação sob controlo**

O Retrievio disponibiliza um conjunto de estatísticas e informações de gestão que refletem o estado do sistema.

Estas informações permitem monitorizar as fontes de informação e consultar o seu estado através de estatísticas, como distribuição do número de registos por sistema recolhido, registos e fontes de informação mais consultados, crescimento mensal, relatórios de agregação, estatísticas de acesso, entre outros.

O sistema de estatísticas do Retrievio é configurável podendo responder mais facilmente às necessidades de quem gere o software.

Qualidade da informação

Para garantir a qualidade da informação recolhida, o Retrievio dispõe de um motor de verificação de conformidade que lhe permitirá validar a informação recolhida de acordo com os critérios estabelecidos. Os critérios de qualidade são totalmente configuráveis e adaptáveis a qualquer situação ou contexto organizacional. Adicionalmente, o sistema pode ser configurado para "limpar" a informação recolhida, produzindo resultados com maior qualidade e mais uniformes.

Rápido como um piscar de olhos

O aumento do número de registos na instância central do Retrievio, resultado da constante incorporação de novas fontes de informação, obrigou a que fossem criadas soluções de engenharia capazes de suportar, eficazmente, grandes volumes de informação.

As agregações são agora mais eficientes pois permitem a recolha de informação em paralelo e as pesquisas mais rápidas e precisas devido ao seu motor de pesquisa de nova geração.

Os seus ativos de informação nunca foram tão valiosos

O Retrievio é o produto ideal para localizar informação no interior da sua instituição.

Recorrendo ao seu elevado número de conetores é possível configurar o Retrievio para pesquisar pastas partilhadas, localizar registos no ERP da sua organização ou encontrar as últimas notícias publicadas no sítio web da sua instituição.

Simple, atraente e intuitivo

A interface gráfica do Retrievio é inteiramente parametrizável podendo ser configurada, através do módulo de administração, para ir ao encontro da identidade gráfica do cliente.

Toda a interface gráfica tem um aspeto simples e intuitivo.

As opções de pesquisa incluem pesquisa básica, pesquisa em texto integral, pesquisa facetada e pesquisa avançada.

API REST documentada através OpenAPI/Swagger

O Retrievio disponibiliza uma API REST documentada em OpenAPI/Swagger que permite a integração com aplicações desenvolvidas por terceiros.

Assim, é possível, por exemplo, desenvolver novas aplicações para acesso ao portal e à informação que este detém, tal como aplicações móveis ou outros serviços de valor acrescentado.

Compatível com qualquer plataforma

O Retrievio é uma solução 100% Web. Para utilizar o sistema apenas necessita de um browser. O software está centralizado num servidor, não havendo a necessidade de instalação de aplicações nos postos de trabalho dos utilizadores. O acesso ao sistema poderá ser feito em qualquer parte e a partir de qualquer dispositivo, *desktop* ou *tablet*.

ARQUITETURA E MÓDULOS APLICACIONAIS

O software Retrieve é composto por 4 módulos aplicativos, os quais estão representados na figura seguinte.

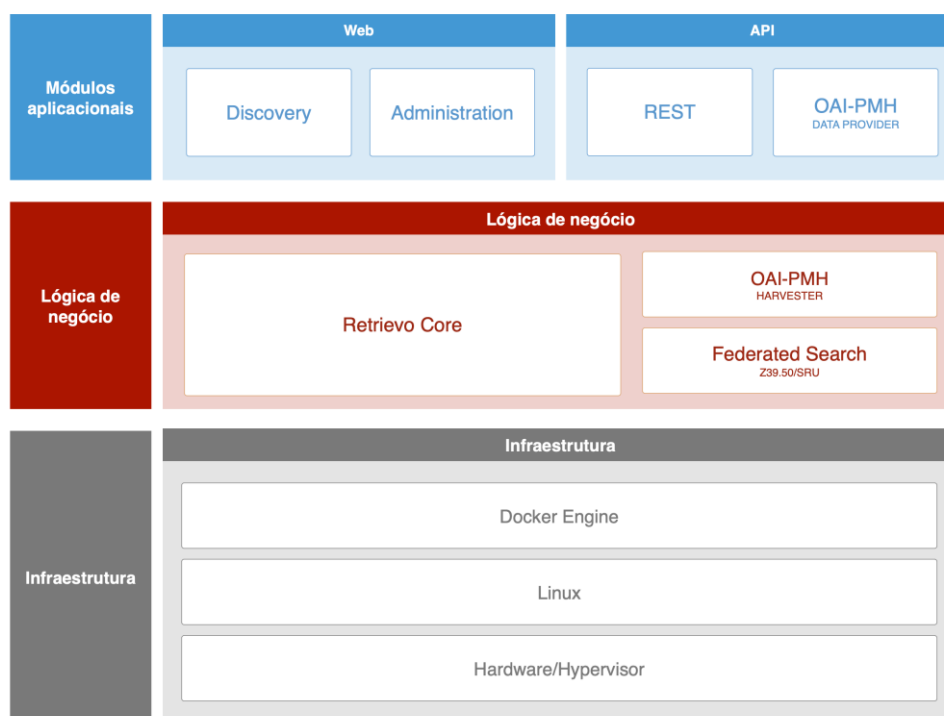


Figura 1 - Arquitetura do Retrieve.

DISCOVERY

O módulo Discovery fornece ao utilizador um ponto de acesso único de pesquisa para recuperação de informação originária de diferentes fontes de informação. O principal objetivo é facilitar o acesso à informação, uma vez que fornece, sobre uma única interface, capacidades de descoberta e recuperação de informação proveniente de fontes de dados heterogéneas.

O utilizador poderá realizar pesquisas transversais em todos os registos recolhidos (pesquisa simples) ou realizar pesquisas mais específicas em determinados campos de informação, sobre determinadas fontes de informação, em conjugação com outros filtros (pesquisa avançada). Pode, por exemplo, pesquisar sobre todos os campos de todos os registos de todos os repositórios, ou simplesmente pesquisar no Título dos repositórios A e B.

Os resultados da pesquisa são consolidados e indexados por título, autor, assunto, datas, ou outros atributos previamente configurados.

Por norma, os resultados de pesquisa apresentam apenas um subconjunto da metainformação do registo original, a qual deverá ser suficiente para identificar a informação que se procura. Para auxiliar a identificação dessa informação, também é recolhida e apresentada, quando existente, uma miniatura do documento associado.

Após localizar a informação pretendida, o utilizador pode consultar o registo de informação no seu contexto original, sendo encaminhado para a fonte de dados que forneceu esse registo, no qual será apresentada a informação completa (metainformação e ficheiros associados, caso existam), e a partir do qual poderão ser desencadeados serviços adicionais (e.g. empréstimo de documentos).

O design gráfico deste módulo pode ser adaptado para ir ao encontro da identidade visual do cliente.

ADMINISTRATION

O Retrieve é acompanhado de um módulo de administração que permite ao utilizador configurar e gerir todo o sistema. Entre as operações disponíveis, é possível listar, modificar e adicionar novas fontes de informação para agregação e disponibilização nas interfaces de pesquisa, bem como configurar os seus parâmetros de agregação.

Permite também consultar estatísticas de agregação, relatórios de problemas ocorridos durante a agregação, estatísticas de crescimento do número de registos por fonte de dados, bem como estatísticas de consulta, entre outros indicadores.

REST API

Para além de suportar todas as funcionalidades dos 2 módulos anteriores, a REST API é uma mais-valia para o cliente pois alarga horizontes no que diz respeito a possíveis integrações com outros sistemas, permitindo o controlo total do Retrieve e sua informação através de interfaces programáticas.

Baseada em standards e nas melhores práticas, a REST API do Retrieve foi documentada usando OpenAPI (anteriormente conhecida por Swagger), disponibilizando online a sua documentação, num formato fácil de ler e de testar.

A REST API do Retrieve é totalmente segura, uma vez que apenas os utilizadores autenticados podem utilizar a API, e deste modo, controlar o produto ou consultar a sua informação.

OAI-PMH DATA PROVIDER

O Retrievio recolhe e concentra informação de várias fontes de informação. Esta informação é validada, limpa, normalizada e devidamente indexada para suportar as pesquisas realizadas através do Módulo Discovery.

Uma vez reunida a informação anteriormente dispersa, o Retrievio pode fornecer informação a outros portais agregadores de conteúdos (e.g. Europeana, Portal Europeu de Arquivos, OpenAire, etc.), recorrendo ao módulo OAI-PMH Data Provider.

Este módulo permite, assim, a criação de um canal normalizado para fornecimento de informação a outros portais, evitando a necessidade de estabelecer adesões a título individual para cada uma das fontes de dados associadas ao Retrievio.

Este módulo disponibiliza a informação através do protocolo OAI-PMH ou através de ficheiros para envio através de mecanismos assíncronos (e.g. FTP).

Criação de **redes de informação**

O Retrievio é perfeito para suportar redes de repositórios de informação num contexto local, nacional ou internacional.

No âmbito local pode ser utilizado, por exemplo, para concentrar e dar acesso a informação de carácter cultural e/ou patrimonial produzida por instituições de um município.

No âmbito nacional ou internacional pode ser utilizado para concentrar informação de diferentes entidades. Exemplos deste tipo de arquitetura são o Repositório Científico de Acesso Aberto de Portugal e o Portal Português de Arquivos.

ACESSIBILIDADE DE CONTEÚDOS WEB

As Diretrizes de Acessibilidade para Conteúdo Web (WCAG) 2.0 são um conjunto de recomendações emanadas pela W3C que visam tornar o conteúdo Web mais acessível. O cumprimento destas diretrizes faz com que o conteúdo publicado na Web se torne acessível a pessoas com incapacidade, nomeadamente cegueira e baixa visão, surdez e baixa audição, dificuldades de aprendizagem, limitações cognitivas, limitações de movimentos, incapacidade de fala, fotossensibilidade, entre outros. Seguir estas diretrizes permite também que o conteúdo Web se torne mais usável por utilizadores em geral e por dispositivos móveis como *smartphones*, *tablets* ou relógios de pulso.

Dada a importância desta temática, foi criada legislação que visa promover a adoção destas diretrizes por todo o Estado português. O Regulamento Nacional de Interoperabilidade Digital (RCM n.º 91/2012 de 8 de novembro) esclarece a obrigatoriedade do cumprimento das diretrizes WCAG 2.0 nível AA por todos os sítios Web do Estado que disponibilizem serviços online¹. O artigo 2º da lei n.º 36/2011 determina que a lei se aplica a:

- Órgãos de soberania;
- Serviços da administração pública central, incluindo institutos públicos e serviços desconcentrados do Estado;
- Serviços da administração pública regional;
- Sector empresarial do Estado.

A KEEP SOLUTIONS é solidária com esta iniciativa e assegura que todos os seus produtos estão em plena conformidade com o nível AA+ das Diretrizes de Acessibilidade para Conteúdo Web (WCAG) 2.0.

¹ Ver também Lei n.º 36/2011, de 21 de junho. Estabelece a adoção de normas abertas nos sistemas informáticos do Estado (PDF, 188KB); RCM n.º 91/2012 de 8 de novembro. Regulamento Nacional de Interoperabilidade Digital (PDF, 16,1MB); DGAEP. (Fevereiro de 2013). Estrutura orgânica da AP Portuguesa pós-PREMAC (PDF, 1,8MB); DGTF. (30/09/2012). Definição de Sector Empresarial do Estado (PDF, 102KB).

REQUISITOS TÉCNICOS

O Retrievo pode ser instalado em qualquer servidor Linux com características de *hardware* convencionais.

Dado que toda a aplicação funciona em ambiente Web, os postos de trabalho dos utilizadores apenas necessitam de um browser para aceder à aplicação (e.g., Internet Explorer, Google Chrome, Mozilla Firefox ou outro).

A tabela seguinte ilustra as características técnicas de um servidor capaz de suportar uma instância do Retrievo com 8 milhões de registos.

SERVIDOR

RAM	8 GB 16 GB recomendados
CPU	Intel Quad-Core ou superior
Disco	100 GB Depende do número de registos, imagens associadas e da sua taxa de crescimento
Sistema operativo	Ubuntu 20.04 LTS ou compatível Sem custos de licenciamento
Software	Docker engine Sem custos de licenciamento
Rede	100 Mbit/s ou superior 1 Gbit/s é recomendável em contexto de incorporação de grandes volumes de representações digitais

POSTO DE TRABALHO

RAM	4 GB
CPU	Intel Dual-Core ou superior
Monitor	1280x768 pixéis ou superior
Sistema operativo	Windows/Linux/MacOS
Software	Web browser
Rede	100 Mbit/s ou superior 1 Gbit/s é recomendável em contexto de incorporação de grandes volumes de representações digitais



www.keep.pt



+351 253 066 735



info@keep.pt



sales@keep.pt



KEEP SOLUTIONS, LDA.
Rua Rosalvo de Almeida, n° 5,
4710-429 Braga
Portugal

KEEP SOLUTIONS

A KEEP SOLUTIONS é uma empresa que tem como missão o fornecimento de soluções avançadas para gestão e preservação de informação.

A nossa abordagem assenta no fornecimento de sistemas de informação e na prestação de serviços com vista a permitir aos nossos clientes realizar uma gestão mais eficiente dos seus ativos de informação.

A empresa iniciou a sua atividade em 2008, tendo adquirido o estatuto de spin-off académica da Universidade do Minho, por se tratar de uma iniciativa empresarial com fortes laços de cooperação com centros de investigação e departamentos desta instituição.

Os nossos clientes encontram-se principalmente no setor público, nos domínios cultural, educacional, patrimonial e científico, mais concretamente nas áreas de arquivo, biblioteca e museu.

Apostamos no desenvolvimento contínuo de soluções inovadoras. Para isso, permanecemos ativos na produção de conhecimento científico, participando ativamente em projetos de I&D em cooperação com instituições nacionais e internacionais.