

Indexação de Documentos Clínicos



Estudante: João Correia

Orientador: Prof. Gabriel David

Proponente: Engº Francisco Correia



Sumário

- Contexto
- Problema
- Objetivo
- Solução
- Revisão Tecnológica
- Implementação
- Arquitetura
- Interface
- Simulação e Testes
- Trabalho Futuro

Contexto

- Numa perspetiva clínica:
 - Existe muita informação, muitos dados
 - Em documentos não estruturados
 - e estruturados em bases de dados
 - É necessário organizar e melhorar o acesso integrado da informação
 - Existem ferramentas de indexação e pesquisa de informação

Problema

- Para quem?
 - Médicos
- Porquê?
 - A informação está:
 - Desorganizada
 - Dispersa por várias fontes
 - Atrasa os médicos
 - Dificulta o acesso à informação



Dados Clínicos

- Resultados analíticos laboratoriais
- Relatórios de imagiologia, cardiologia, etc...
- Diagnósticos codificados por exemplo em ICD9
- Notas clínicas dos médicos
- Requisições de exames
- Prescrições de medicamentos
- Informação demográfica de pacientes

Objetivo

- O objetivo da dissertação é:
 - Permitir uma pesquisa em texto integral que combine conteúdo de documentos com conteúdo de colunas de base de dados
 - Aglomerar a informação usando o openEHR
 - Apresentar os resultados tipo Google

Solução proposta

- Converter os dados da BD fonte (EResults da Glintt) para openEHR
- Indexar com o Solr
- Preparar uma interface de pesquisa

Revisão de Tecnologias

Converter os dados da BD
fonte (EResults da Glintt)
para openEHR

Indexar com o Solr

Preparar uma interface de
pesquisa



EHRServer



- Repositório de dados clínicos
- Segue a norma openEHR
- Open Source
- REST API

Converter os dados da BD fonte (EResults da Glintt) para openEHR

Indexar com o Solr

Preparar uma interface de pesquisa

Norma openEHR

- Norma utilizada para registos clínicos digitais
- A norma openEHR está estruturada da seguinte forma:
 - **Arquétipo:** conceito mais elementar nesta estrutura (equivalente a uma peça de LEGO), consiste em dados clínicos, como por exemplo, altura, peso, sumário de gravidez e ecocardiograma;
 - **Template:** consiste num conjunto de peças LEGO, isto é, num conjunto de arquétipos.

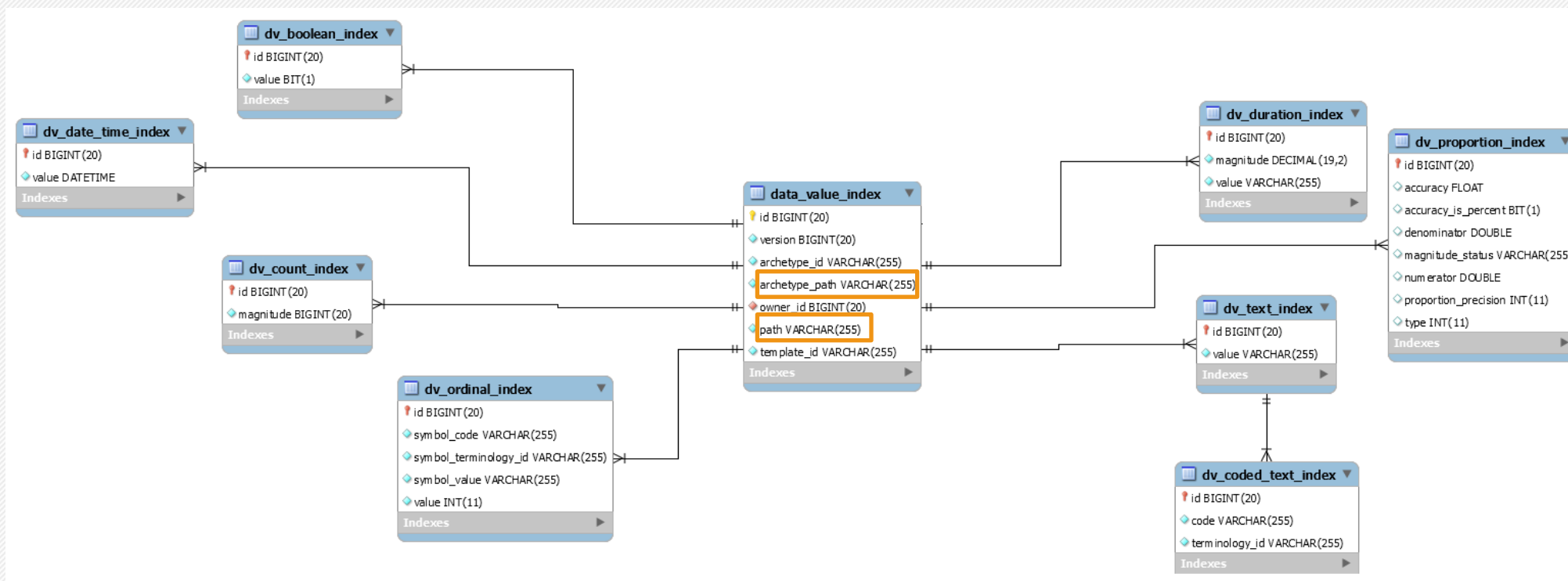
Converter os dados da BD fonte (EResults da Glintt) para openEHR

Indexar com o Solr

Preparar uma interface de pesquisa

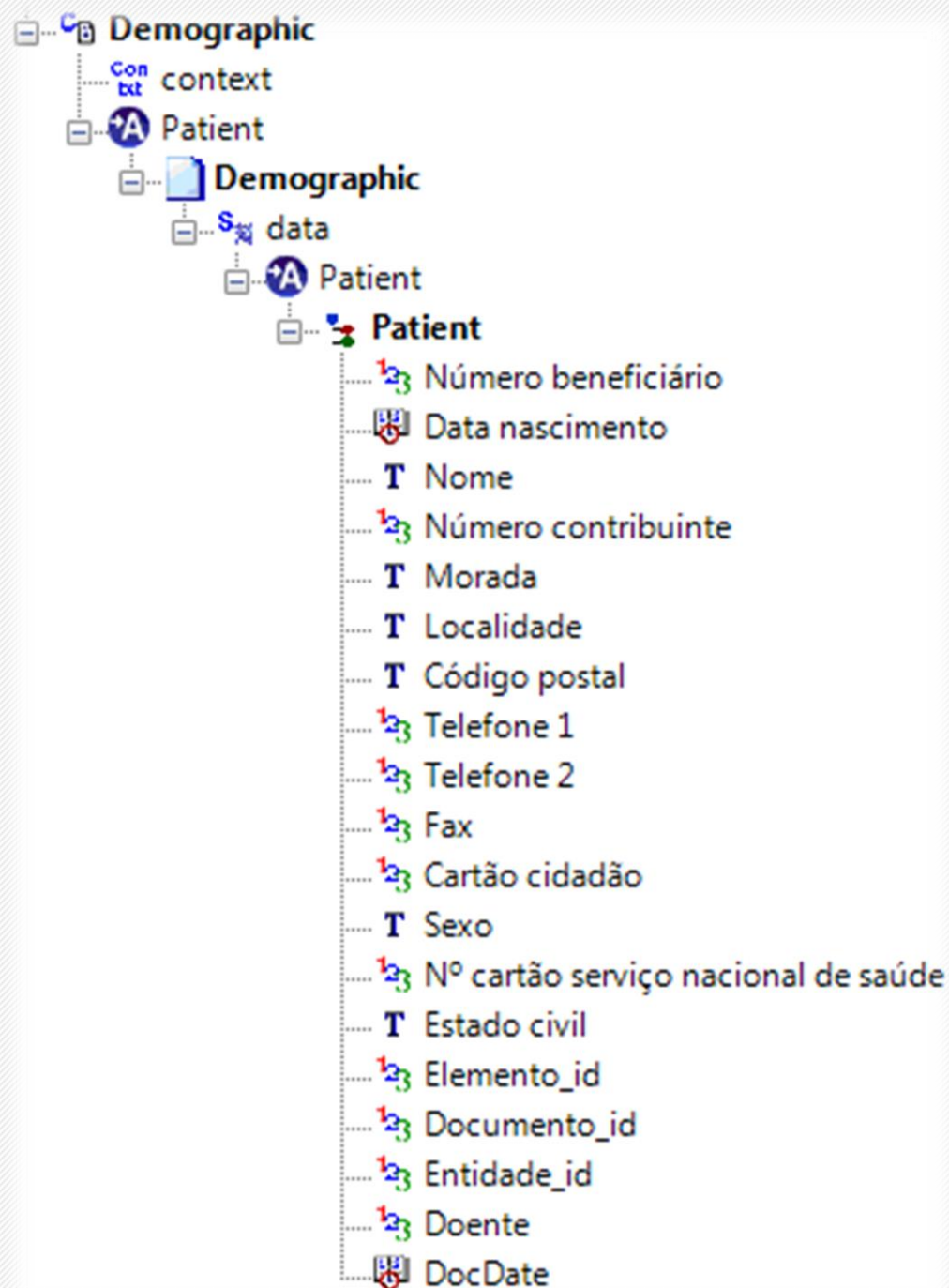
Estrutura EHRServer

Converter os dados da BD fonte (EResults da Glintt) para openEHR



Indexar com o Solr

Preparar uma interface de pesquisa



Converter os dados da BD fonte (EResults da Glintt) para openEHR

Indexar com o Solr

Preparar uma interface de pesquisa

Modelo de Referência

Converter os dados da BD fonte (EResults da Glintt) para openEHR

EHR	Registo clínico eletrónico por pessoa
Folders	Organização de alto nível do EHR (ex.: por episódio, por especialidade clínica)
Compositions	Conjunto de entradas (Entries) submetidas numa determinada data (ex.: relatório)
Sections	Cabeçalhos clínicos que refletem o fluxo de trabalho
Entries	Declarações clínicas sobre observações, avaliações e instruções
Clusters	Entradas compostas
Elements	Entradas elementares
Data values	Termos codificados do conjunto de termos

Indexar com o Solr

Preparar uma interface de pesquisa

Indexar com o Solr



- solrconfig.xml
 - Inclui as bibliotecas necessárias
- schema.xml
 - Estrutura dos dados a serem tratados pelo Solr
 - Definição dos parâmetros de indexação e pesquisa

Converter os dados da BD fonte (EResults da Glintt) para openEHR

Indexar com o Solr

Preparar uma interface de pesquisa

Pesquisar com o Solr

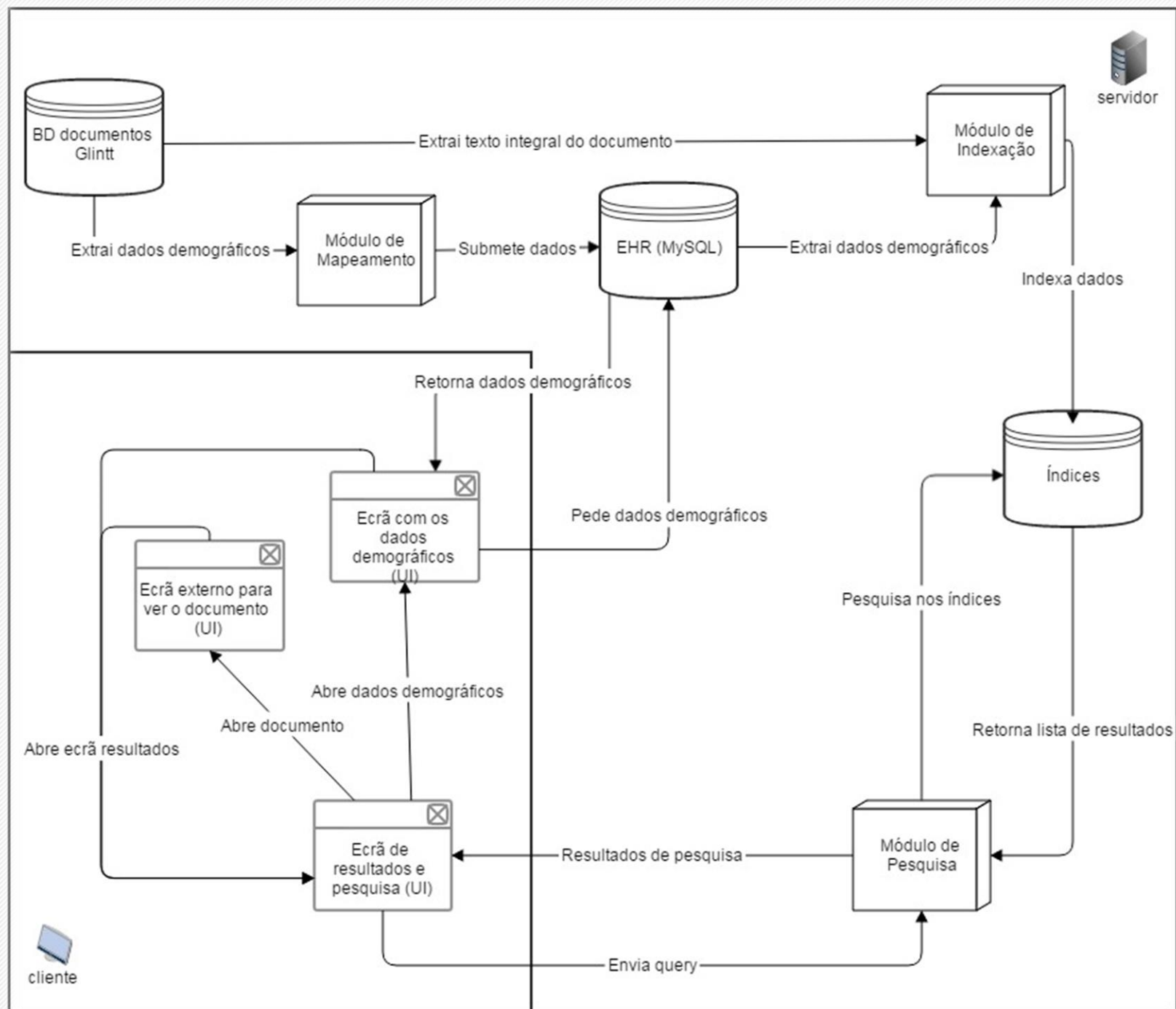
- Por ordem de relevância
 - Modelo Vetorial
 - Modelo Booleano

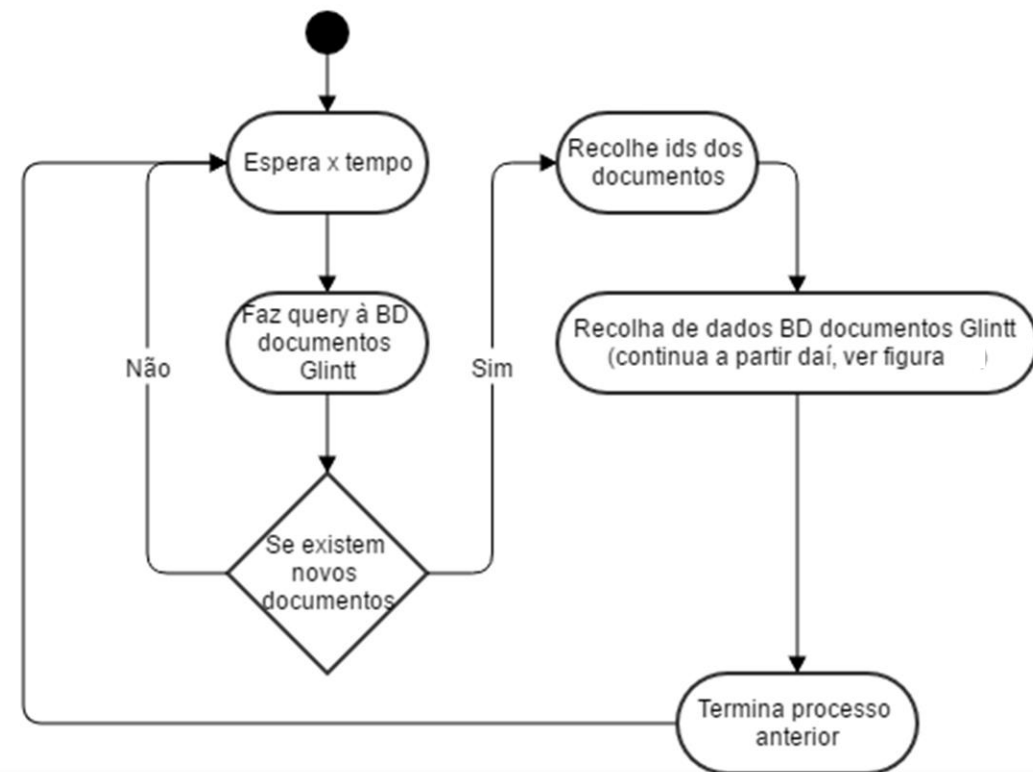
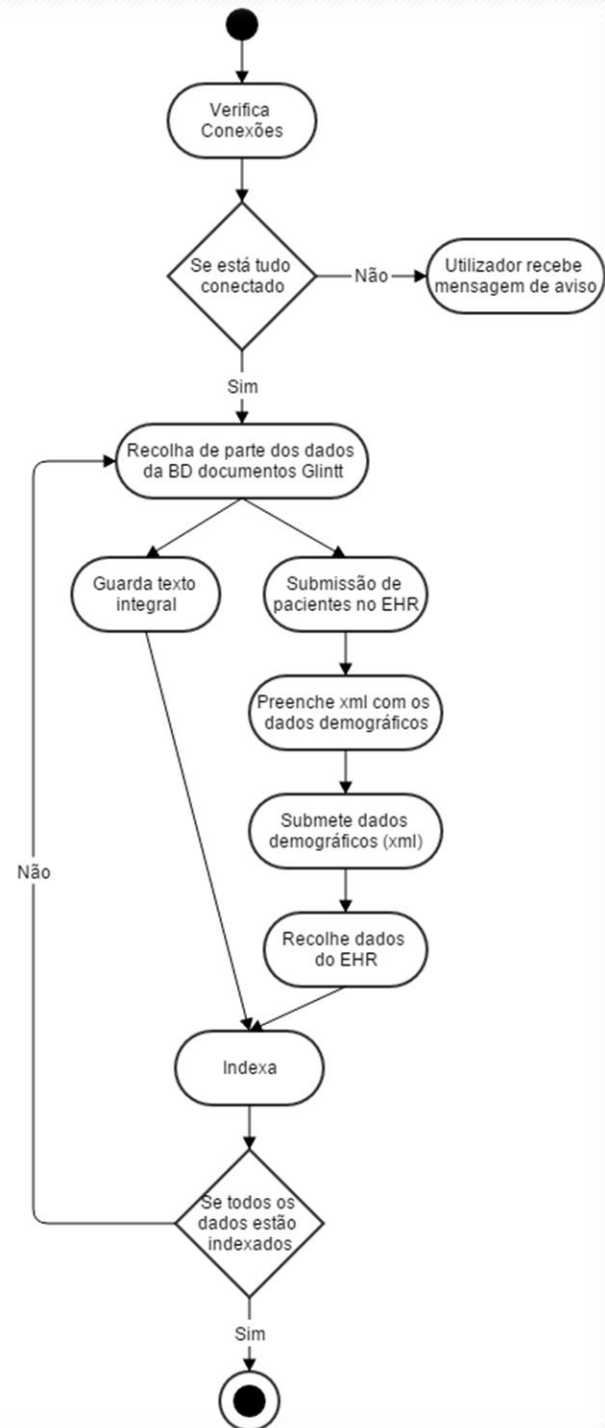
Converter os dados da BD fonte (EResults da Glintt) para openEHR

Indexar com o Solr

Preparar uma interface de pesquisa

Arquitetura





Interface

Indexação de documentos clínicos

ProcurarPesquisa Avançada

██

ERITRÓCITOS 4,97 106 / μ L 3.50 - 5.50 HEMOGLOBINA 14,0 g/dL 12.0-15.0 HEMATÓCRITO 42,5 % 39.0 - 55.0 VGM 85,5... NormalCORPOS CETÓNICOS NegativoPIGMENTOS BILIARES NegativoUROBILINOGÊNIO 0,2 mg/dl HEMOGLOBINA 0,10 mg/dl

[Documento](#) - [Informação Demográfica](#)

██

ERITRÓCITOS 4,58 106 / μ L 3.50 - 5.50 HEMOGLOBINA 13,8 g/dL 12.0-15.0 HEMATÓCRITO 41,4 % 39.0 - 55.0 VGM 90,4... NormalCORPOS CETÓNICOS NegativoPIGMENTOS BILIARES NegativoUROBILINOGÊNIO 0,2 mg/dl HEMOGLOBINA 0,20 mg/dl

[Documento](#) - [Informação Demográfica](#)

██

ERITRÓCITOS 4,03 106 / μ L 3.50 - 5.50 HEMOGLOBINA 12,1 g/dL 12.0-15.0 HEMATÓCRITO 37,0 % 39.0 - 55.0 VGM 91,6

[Documento](#) - [Informação Demográfica](#)

██0

Informaticamente por : HEMOGRAMA ERITROGRAMA ERITRÓCITOS 3,77 106 / μ L 3.50 - 5.50 4,00 3,94 3,94 HEMOGLOBINA

[Documento](#) - [Informação Demográfica](#)



Indexação de documentos clínicos

[Procurar](#)[Pesquisa Avançada](#)

De

14 de Junho de 2016



a

23 de Junho de 2016



Simulação e Testes

	Máquina 1	Máquina 2
Processador	Genuine Intel® CPU U7300 @1.30GHz	Intel® Core™ i7-4510U CPU @ 2.00GHz
RAM	4GB	8GB
Sistema Operativo	Windows 7 64bits	Windows 10 64 bits

Simulação e Testes

	Máquina 1			Máquina 2		
Tamanho da partição	100	500	1000	100	500	1000
Processo de Indexação	1min20seg	33seg	25seg	40seg	16seg	11seg
Processo de Mapeamento	3h50min	2h04min	3h37min	1h17min	56min	1h13min
Todos os processos	3h52min	2h05min	3h38min	1h18min	57min	1h14min

Número total de documentos: **3680**

Tamanho médio dos documentos: **161KB**

Formato dos documentos: **txt, pdf, jpeg, png, docx, xml**

Conclusões e Satisfação dos Objetivos

Trabalho Futuro

- Pesquisa por hierarquia de doenças e por região
- Dicionários de sinónimos e *spellcheck*
- Dividir o template existente em outros, para poder ser reaproveitado

