# Indexação e Tratamento de Dados Heterogêneos: Variedade

Mecanismos de Busca

## Aula 8

Solr & Elastic

Resultados do Trabalho em grupo



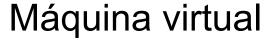
# Máquina virtual

TUTO IVENTE I 1994 PARTIE I 19

Máquina virtual com Elastic, Solr e Nutch (webcrawler) já instalados

Usuário: infnet

Senha: infnet





Solr pode ser inicializado navegando-se até a pasta do Solr e digitando

bin/solr start (máquina virtual já vem com os dois serviços inicializados)

Solr (versão tomcat): <a href="http://localhost:8080/solr/">http://localhost:8080/solr/</a>

Solr (versão jetty): <a href="http://localhost:8983/solr/">http://localhost:8983/solr/</a>

Elastic pode ser acessado em: <a href="http://localhost:9200/">http://localhost:9200/</a>

### Solr



Pastas e arquivos do Solr podem ser acessados em:

/usr/share/solr/

Pasta conf contém arquivos de configuração, entre eles o schema.xml

schema.xml é onde se configura que tipos de campos existirão, quais campos deverão ser usados como chave primária, quais os campos que devem ser obrigatórios, e como indexar e buscar cada campo

### Solr



### Principais campos do schema.xml

- fieldType
  - texto, string, data e long/int
- fields
  - o name, type, indexed, stored, compressed...
- UniqueKey entre outros

https://wiki.apache.org/solr/SchemaXml





Solr permite que um banco de dados seja exportado para dentro dele

https://wiki.apache.org/solr/DIHQuickStart

Exemplo de indexação de documentos:

- Navegue até o diretório do Solr (jetty) e execute o comando de indexação
  - pasta: /opt/solr/solr
  - o comando: bin/post -c gettingstarted example/exampledocs/\*.xml

Isso indexará todos os documentos .xml contidos na pasta exampledocs

### Conectando ao Solr

Maiores detalhes podem ser encontrados em <a href="https://lucene.apache.org/solr/5\_4\_1/quickstart.html">https://lucene.apache.org/solr/5\_4\_1/quickstart.html</a>



# Trabalho em grupo

Resultados do trabalho em grupo



### **Nota final**

Notas dos trabalhos entregues já estão disponíveis para quem quiser saber

Nota final (trabalho em grupo + trabalhos semanais) será dada até sexta-feira que vem (23/12/2016)

Trabalhos em grupo atrasados poderão ser entregues até as 18:00 da quinta que vem (22/12/16), depois disso não serão corrigidos!

Envie o trabalho ainda que este não esteja completo, pois com certeza valerá algo

Notas serão enviadas por email para cada aluno

