



Universidade Federal de Ouro Preto

UFOP ATIVA: UMA FERRAMENTA PARA EXTRAÇÃO E ANÁLISE DE DADOS ACADÊMICOS DE SERVIDORES DA UFOP BASEADA NO CURRÍCULO LATTES

Aluno: Mateus Vitor Pereira Lana

Orientador: Anderson Almeida Ferreira

Co-orientador: Helgem de Souza Ribeiro Martins

Ouro Preto, 2019



Organização da Apresentação

- Introdução
 - Contextualização;
 - Justificativa;
 - Objetivos.
- Trabalhos Relacionados
- Desenvolvimento
 - Modelagem do Banco de Dados;
 - Concepção do Software;
 - Extração dos Dados;
 - Mapeamento dos Dados;
 - Deduplicação de Produções;
 - Implementação do Sistema.
- Estudo de Caso
- Conclusão e Trabalhos Futuros

INTRODUÇÃO





Contextualização

- A pesquisa faz parte do processo de desenvolvimento de um indivíduo. Portanto, ela pode ser considerada uma das atividades mais importantes desenvolvidas ao longo da vida acadêmica;
- O surgimento da pesquisa científica trouxe consigo progressos de conhecimentos em diversas áreas, convertendo-se em vantagens sociais e econômicas.



Contextualização

- Nos últimos quarenta anos, a pesquisa científica no nosso país deu um salto significativo.
 - Política de pós-graduação implantada nos anos sessenta.
- Levantamentos realizados pela *Clarivate Analytics* mostram que:
 - Brasil, 2011-2016, 13º produtor mundial de publicações de pesquisa (250 mil *papers*);
 - A produção cresce anualmente.

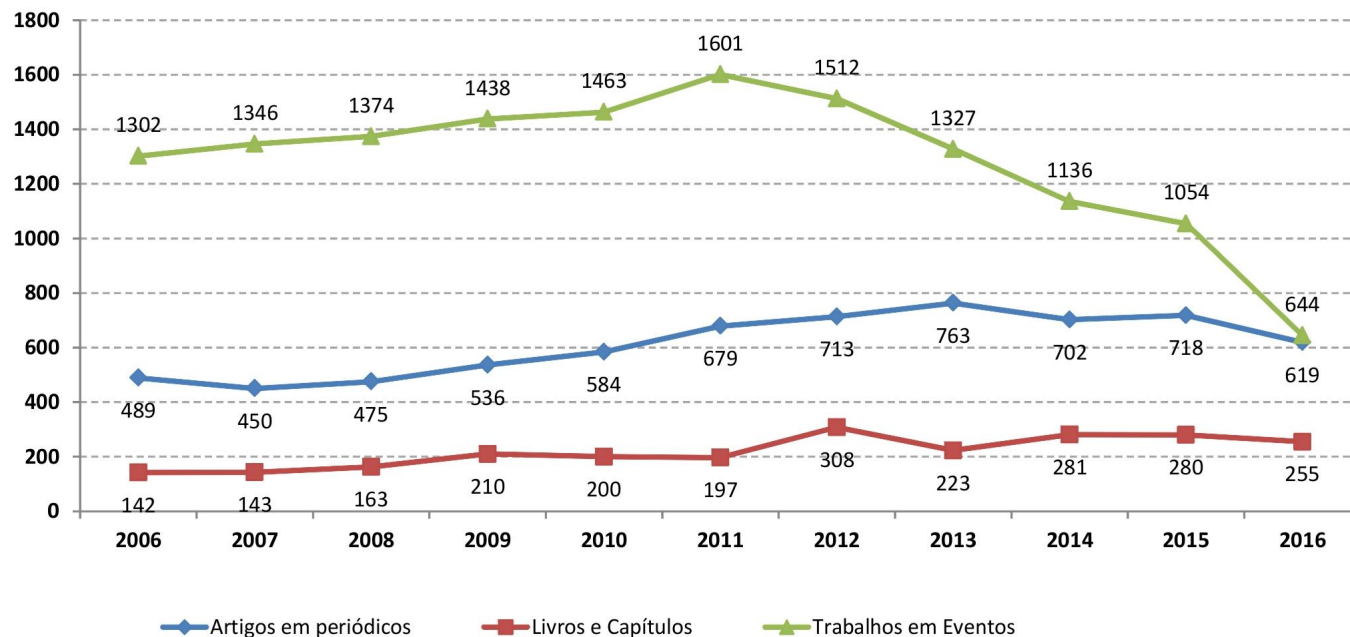


Contextualização

- Atualmente a UFOP conta com:
 - Mais de 1600 servidores;
 - Mais de 900 docentes;
 - 55 cursos;
- Grande parte destes docentes, parte dos alunos (principalmente de pós-graduação) e também técnicos, estão envolvidos em diversos projetos, realizam pesquisas, orientações, etc.

Contextualização

Evolução das publicações Acadêmicas de Docentes da UFOP




Fonte: Somos UFOP
Elaboração: Alberto Fonseca

Contextualização - Plataforma Lattes

- Desenvolvida pelo Conselho Nacional de Desenvolvimento Científico e Tecnológico(CNPq) ;
- Repositório de dados com currículos, grupos de pesquisa e instituições;
- Criada para facilitar as ações de planejamento, gestão e operacionalização do fomento à pesquisa.





Contextualização - Currículo Lattes

- Apresentação;
- Dados Gerais;
- Produção Bibliográfica;
- Produção Técnica;
- Produção Cultural;
- Orientações;
- Eventos;
- Bancas;
- Indicadores.



Dados semi estruturados

- Dados semi estruturados são dados nos quais o esquema de representação está presente (de forma explícita ou implícita) juntamente com o dado, ou seja, o mesmo é auto-descritivo;
- Isto significa que uma análise do dado deve ser feita para que a sua estrutura possa ser identificada e extraída.

```

<!-- SBB DTD versão 1.3 -->

<!ELEMENT SBB(artigo+)>
<!ELEMENT artigo (título, autor+,
resumo, seção+)>
<!ELEMENT autor (nomea,
instituição+, email*)>
<!ELEMENT seção (nomes, texto)>
<!ELEMENT instituição (nomei,
endereço?)>
<!ELEMENT título (#PCDATA)>
<!ELEMENT resumo (#PCDATA)>
...
<!ATTLIST SBB ano CDATA #REQUIRED
local CDATA "Rio">

```

(a)

```

<?xml version="1.0" encoding="UTF-8"
standalone="no" ?>
<!DOCTYPE SBB SYSTEM SBB.dtd>
<SBB ano="2000">
</artigo>
  <título> Dados Semi-Estruturados </título>
  <autor>
    <nomea> Ronaldo Mello </nomea>
    <instituição>
      <nomei> UFRGS </nomei>
      <endereço> Bento Gonçalves, 9500
      </endereço>
    </instituição>
    <instituição>
      <nomei> UFSC </nomei>
    </instituição>
    <email> ronaldo@inf.ufrgs.br </email>
    <email> ronaldo@inf.ufsc.br </email>
  </autor>
  <autor>
    <nomea> Carlos Heuser </nomea>
    <instituição>
      <nomei> UFRGS </nomei>
      <endereço> Bento Gonçalves, 9500
      </endereço>
    </instituição>
    <email> heuser@inf.ufrgs.br </email>
  </autor>
  <resumo> ... </resumo>
  <seção>
    <nomes> Introdução </nomes>
    <texto> ... </texto>
  </seção>
  ...
</artigo>
...
</SBB>

```

(b)




Justificativa

- Para as instituições de ensino superior como a UFOP, é extremamente importante levantamentos de informações em relação a produção de pesquisa científica realizada pelos diversos servidores, departamentos e unidades.
 - Possibilita transparecer esses dados que são de grande importância internamente;
 - Permite um acesso externo que também contribui para a divulgação da instituição.



Objetivo Geral

- Este projeto teve como principal finalidade o desenvolvimento de um sistema que agrupe e organize em um banco de dados as informações, provenientes da Plataforma Lattes, que estejam relacionadas aos servidores atuantes na UFOP.



Objetivos Específicos

- Contribuir para a facilidade de acesso a informações a respeito dos servidores da UFOP;
- Colaborar com as futuras pesquisas que possam usufruir da plataforma concebida;
- Contribuir para uma maior organização do acervo de informações que estão relacionadas às pesquisas realizadas na UFOP.

TRABALHOS RELACIONADOS





Plataforma Somos (2017)

- Desenvolvida por uma equipe da UFMG;
- Por meio de dados provenientes da Plataforma Lattes;
- Busca facilitar o mapeamento das competências da UFMG e demais instituições de ensino que usufruem do serviço;
- Incrementar a interação dessas universidades em áreas de pesquisa científica e tecnológica com instituições públicas e privadas.

Plataforma Somos (2017)

Última atualização do sistema: 25.06.2019

HOME | INDICADORES | CONTATO | SOBRE

Procurar

Somos UFMG

Professor



Adriano Alonso Veloso <http://lattes.cnpq.br/9973021912226739>

Bolsista de Produtividade do CNPq Nível 2

Última atualização do Lattes: 12.12.2018

UNIDADE: Instituto de Ciências Exatas

DEPARTAMENTO: Departamento de Ciência da Computação

NOMES DE CITAÇÃO: VELOSO, A. A. / Veloso, Adriano / Veloso, A. / VELOSO, ADRIANO ALONSO

Exibir Gráficos

Produção Bibliográfica

NUMERO DE CITAÇÕES: 2004

FONTE: OUTRAS

NUMERO DE CITAÇÕES: 124
FATOR H: 7

FONTE: WEB OF SCIENCE

Artigos Publicados (22, 73% com DOI)

Livros (1, 0% com DOI)

Trabalho em Eventos (78, 44% com DOI)

Propriedade Intelectual

Produto Tecnológico (2)

Software (3)

Plataforma Somos (2017)

Última atualização do sistema: 31.10.2019

[HOME](#) | [INDICADORES](#) | [CONTATO](#) | [SOBRE](#)



Indicadores

NÚMERO DE PROFESSORES DA UFMG	3.175
NÚMERO DE PROFESSORES COM LATTES	3.163 (99,62%)
NÚMERO DE UNIDADES CADASTRADAS	23
NÚMERO DE PATENTES	1042

Atenção: Nestes indicadores não estão contabilizados professores visitantes, substitutos e aposentados. Além disso, 12 professores não possuem Lattes ou o Lattes ainda não foi localizado. Estes professores não foram contabilizados nas estatísticas que envolvem dados do Lattes.

Indicadores das unidades

#	Unidade	# de professores	% do total de professores	# de patentes
01	Faculdade de Medicina	364	11.5%	24
02	Instituto de Ciências Exatas	343	10.8%	322
03	Escola de Engenharia	308	9.7%	244
04	Instituto de Ciências Biológicas	286	9.0%	353
05	Faculdade de Filosofia e Ciências Humanas	213	6.7%	0
06	Faculdade de Letras	149	4.7%	0
07	Faculdade de Educação	145	4.6%	0
08	Faculdade de Odontologia	125	3.9%	23
09	Faculdade de Direito	116	3.7%	0
10	Faculdade de Ciências Econômicas	116	3.7%	0
11	Escola de Belas Artes	111	3.5%	2
12	Escola de Veterinária	110	3.5%	40
13	Escola de Enfermagem	102	3.2%	1
14	Instituto de Ciências Agrárias	94	3.0%	1
15	Escola de Educação Física, Fisioterapia e Terapia Ocupacional	93	2.9%	0
16	Escola de Arquitetura	89	2.8%	7
17	Instituto de Geociências	85	2.7%	0
18	Faculdade de Farmácia	78	2.5%	110
19	Escola de Música	73	2.3%	0
20	Centro Pedagógico	65	2.0%	0
21	Coltec	59	1.9%	52
22	Escola de Ciência da Informação	44	1.4%	0
23	Teatro Universitário	7	0.2%	0

Plataforma Somos (2017)

Última atualização do sistema: 31.10.2019

HOME INDICADORES CONTATO SOBRE

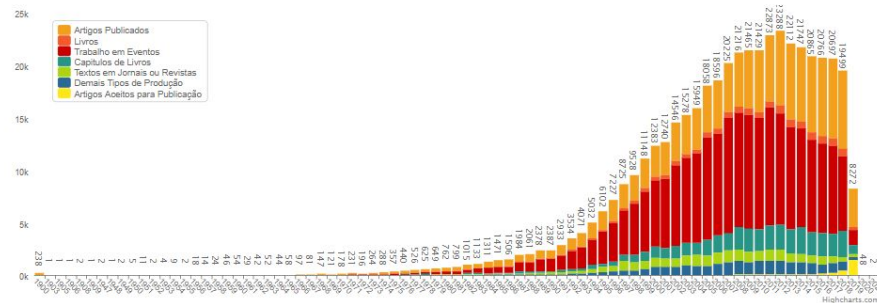


Procurar

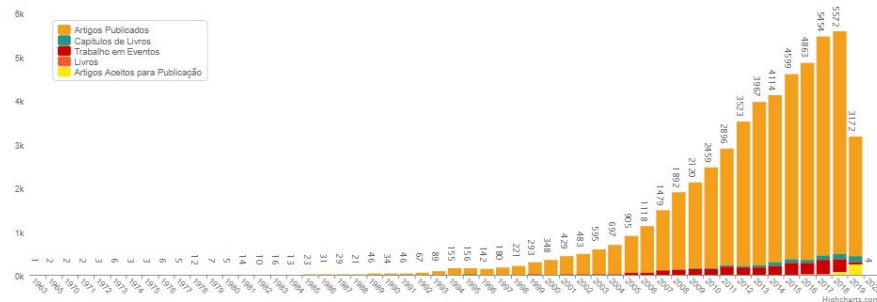
0.2%

Gráficos

Produção Bibliográfica Histórica



Produção Bibliográfica Histórica com DOI



DESENVOLVIMENTO





Tecnologias

Java Architecture
for XML Binding



spring

by Pivotal™



ANGULAR

Maven™



Java™

JPA

Java Persistence API



HIBERNATE

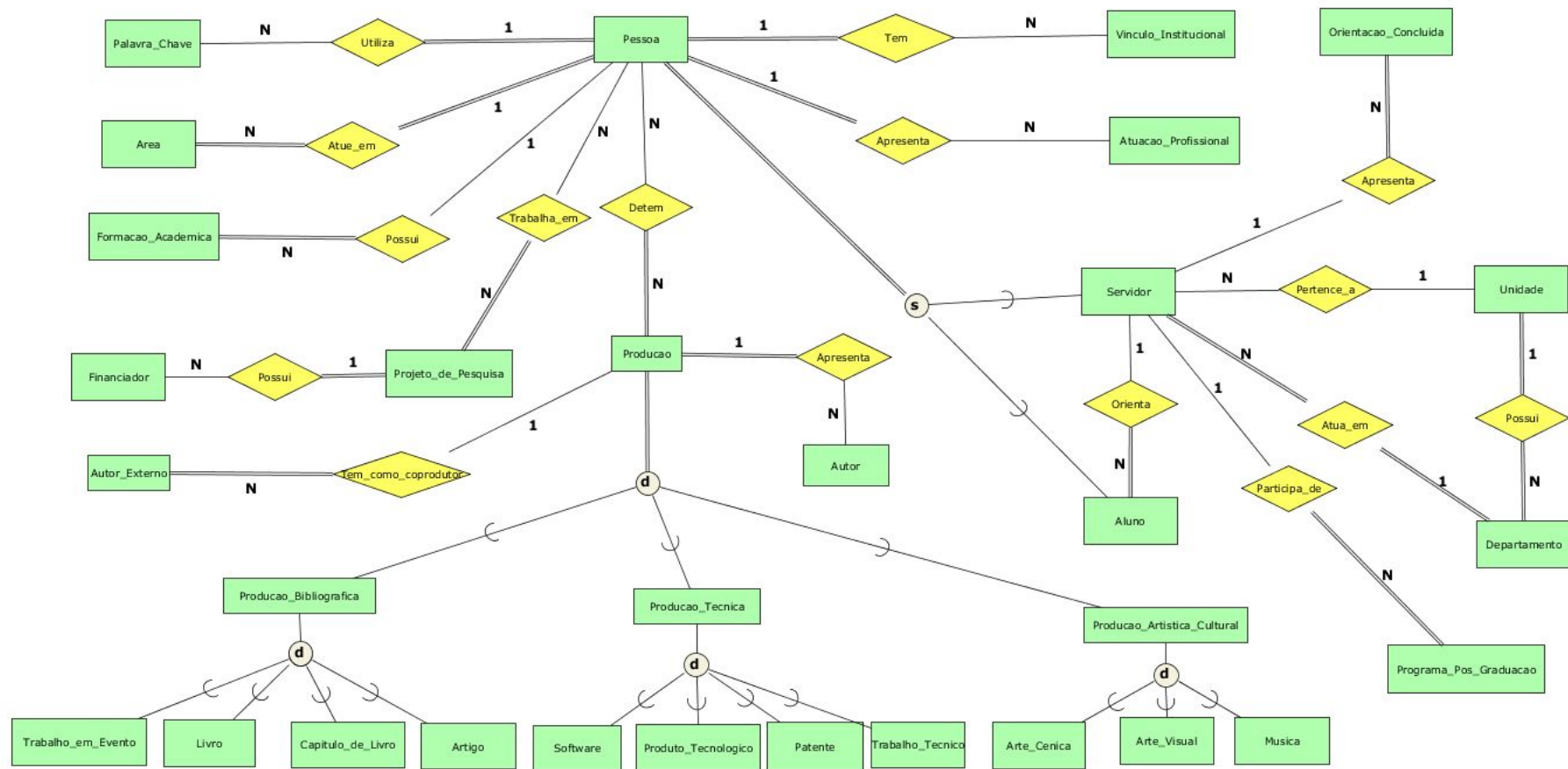


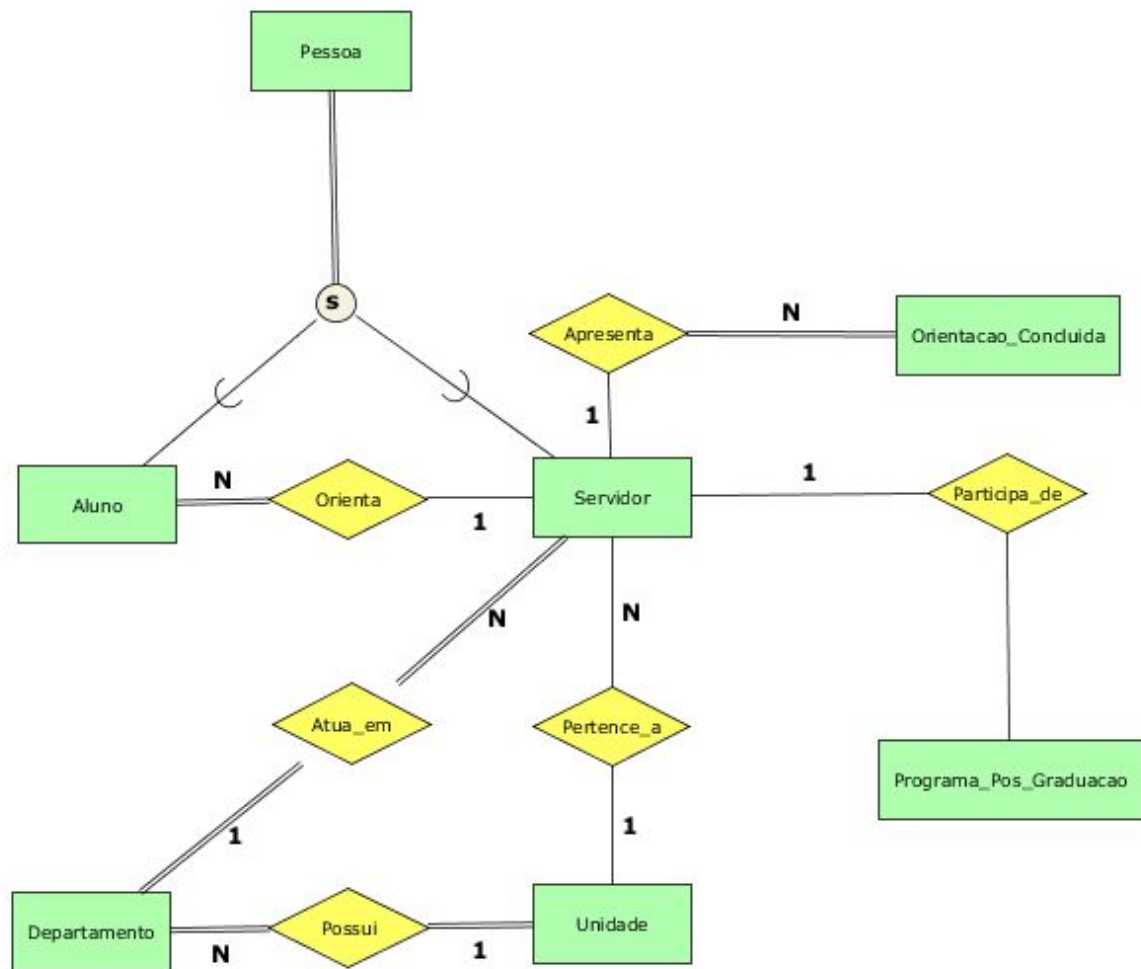
PostgreSQL

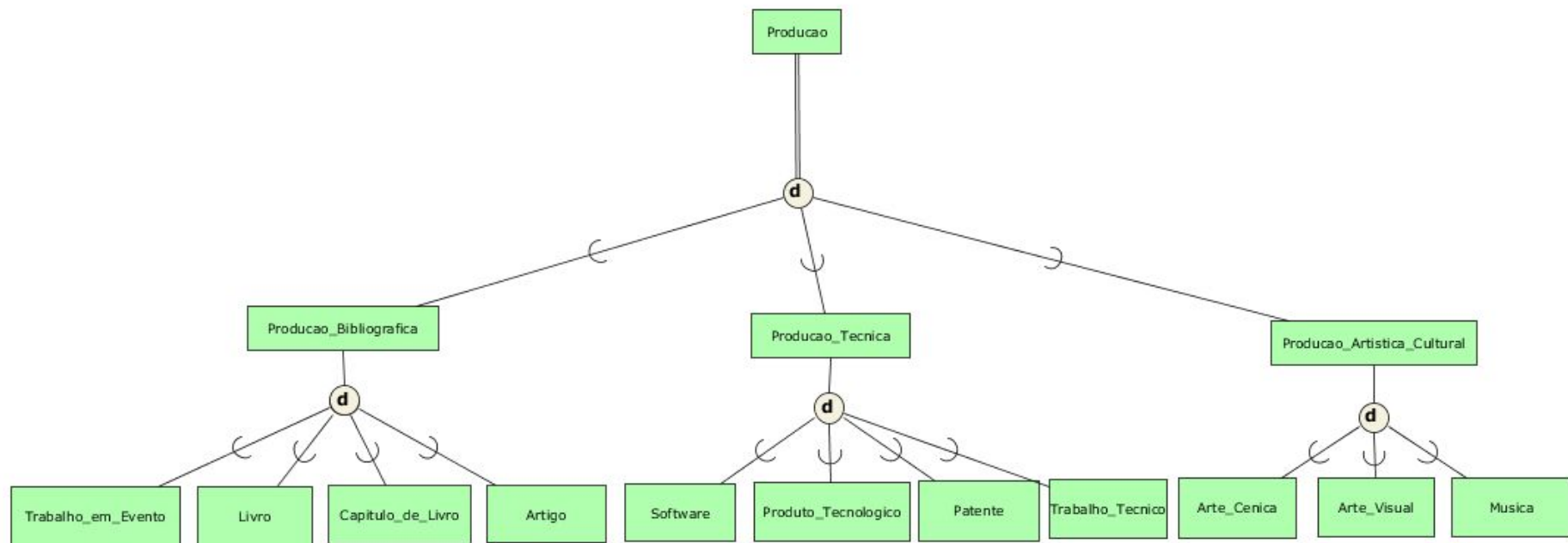


Esquema Conceitual do Banco de Dados

- 31 entidades;
- Mais de 100 atributos;
- 16 relacionamentos.









Processo de Concepção de Software

"[...] é uma abordagem adaptável que possibilita às pessoas (a equipe de software) realizar o trabalho de selecionar e escolher o conjunto apropriado de ações e tarefas [...]"(Pressman e Maxim, 2016).

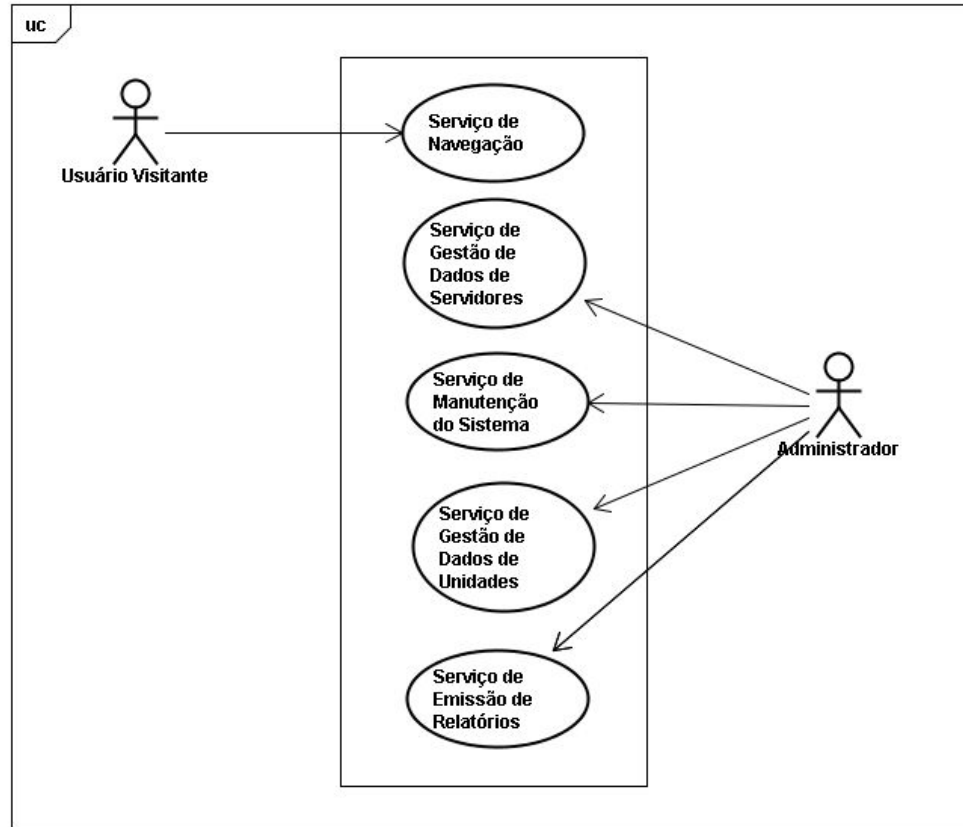
- Documento de Especificação;
- Participação do Cliente;
- Ciclo de vida de desenvolvimento.



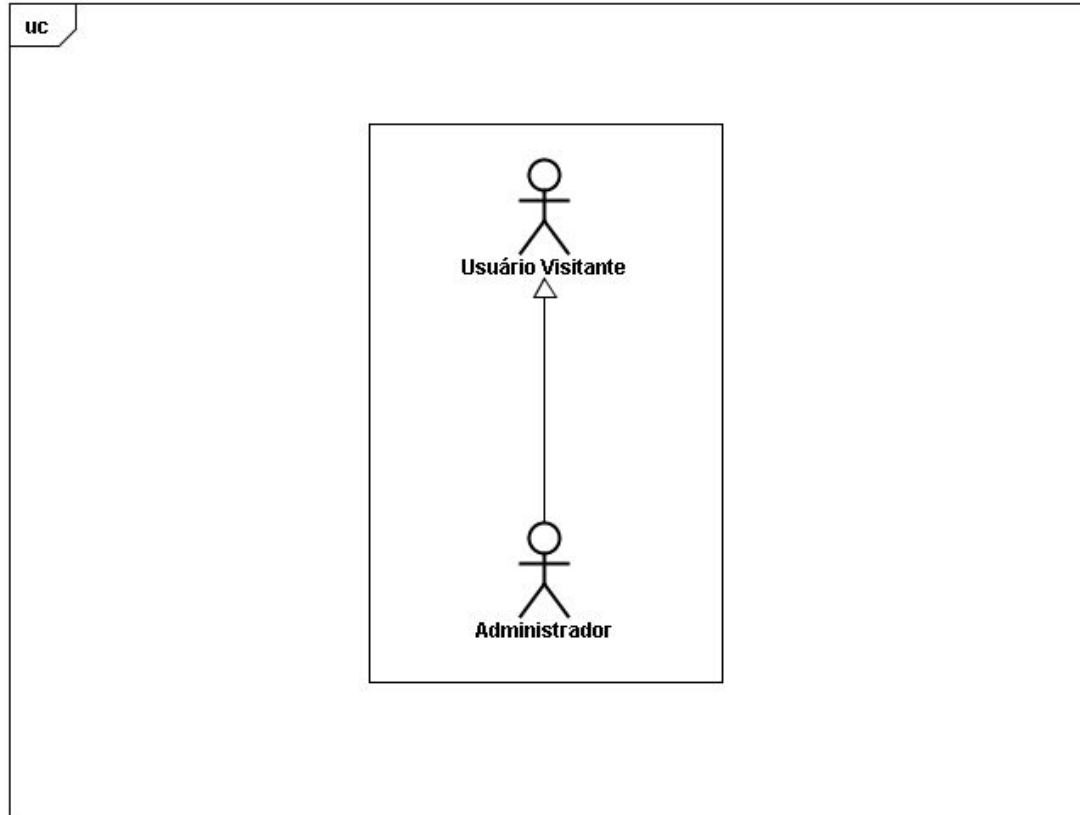
Concepção do Software

- Nome do Produto e Componentes;
- Missão do Produto;
- Descrição do Produto;
- Modelo de Negócio;
- Escopo positivo e negativo.

Concepção do Software - Diagrama de Contexto



Concepção do Software - Hierarquia de Atores





Concepção do Software - Backlog

- Serviço de Navegação:
 1. Efetuar Consulta
 2. Visualizar Dados Relevantes
 3. Visualizar formas de entrar em contato
 4. Visualizar informações sobre a plataforma



Concepção do Software

Estória de Usuário

COMO UM usuário **COM O OBJETIVO** de abrir uma página para ver os dados gerais referentes a UFOP **EU GOSTARIA DE** consultar os dados relevantes.

Cenário de Teste

DADO que eu sou um visitante **E** estou na tela inicial do sistema **E** desejo conferir os dados relevantes referentes a pesquisa realizada na universidade **QUANDO** eu clicar na aba “Dados Relevantes” na parte superior da tela **ENTÃO** uma nova tela com os dados será exibida.



http://ufop.br/lorem

Q

UFOP Ativa

[Home](#)

[Dados Relevantes](#)

[Fale Conosco](#)

[Sobre a Plataforma](#)

Q Lorem

Lorem Ipsum

Lorem Dolor

Lorem Amet

Be

Lorem ipsum dolor sit amet

dui ligula. Pellentesque augue turpis, dapibus eu turpis nec, tristique pellentesque nisi. Mauris finibus lectus sed enim luctus interdum. Vestibulum imperdiet risus at vehicula tempor.

Phasellus bibendum maximus vehicula. In in urna suscipit, gravida ex vel, dictum nunc. Duis egestas tempor arcu. Nulla ac ultrices nibh. Donec consequat blandit lectus, efficitur aliquam elit ullamcorper non. Praesent quis suscipit dolor, non euismod neque. Nunc turpis nunc, volutpat nec molestie a, rutrum vel ligula.



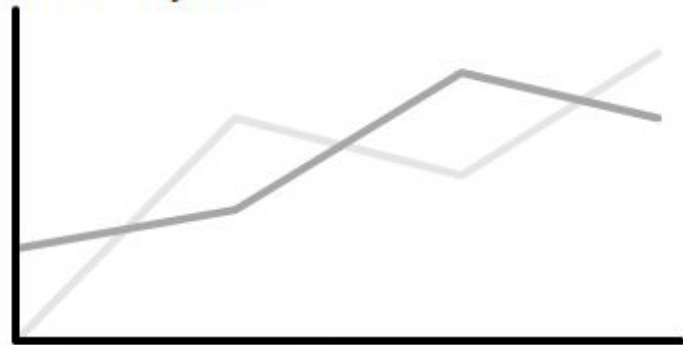


Dados Relevantes

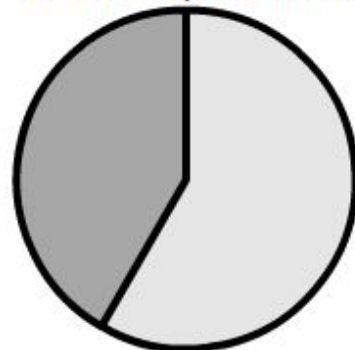
[Home](#)
[Dados Relevantes](#)
[Fale Conosco](#)
[Sobre a Plataforma](#)

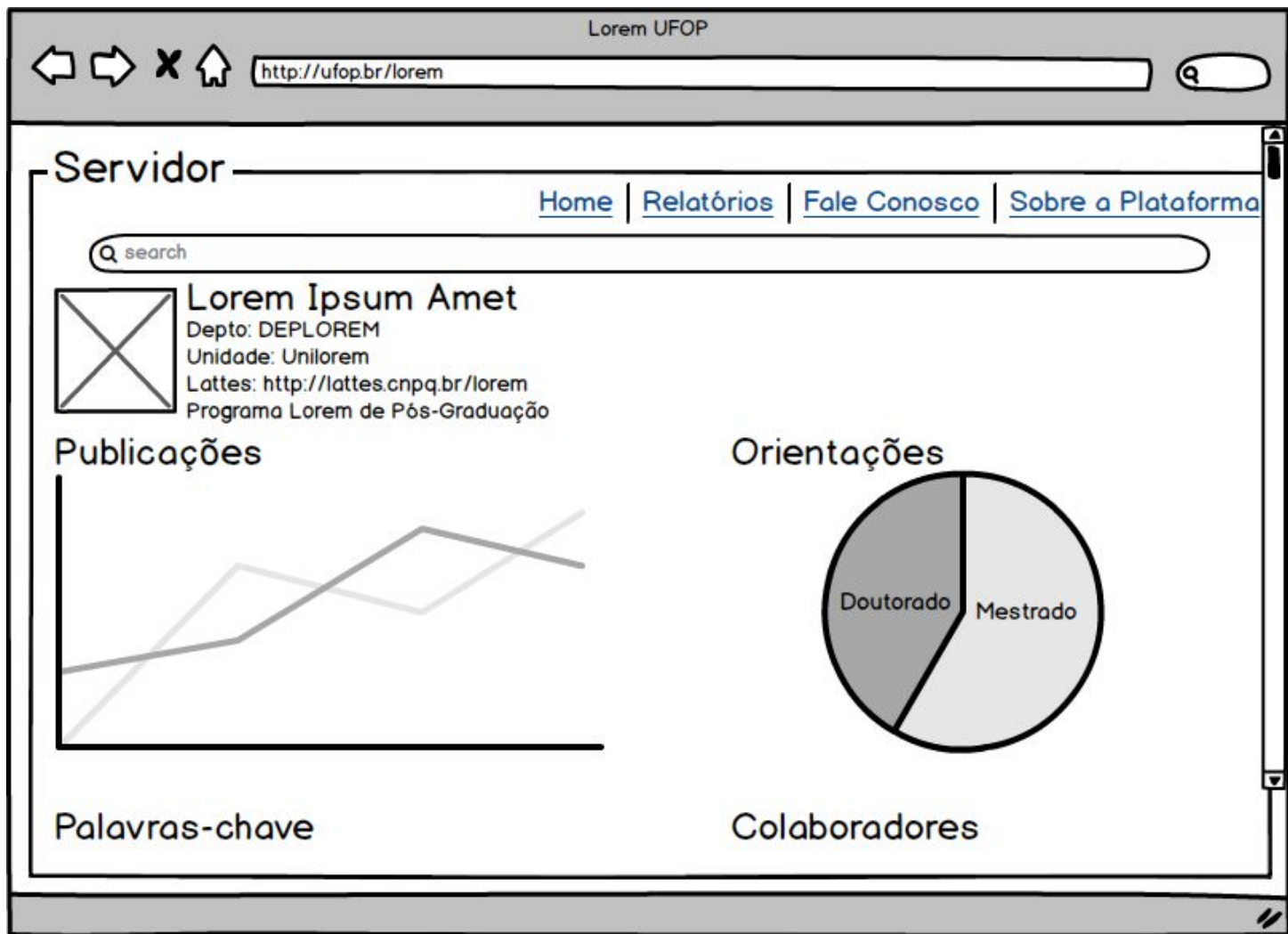
Número de servidores efetivos	900
Número de servidores efetivos com Lattes	878
Números de Unidades	15
Número de Departamentos	10

Publicações



Servidores por unidade







UFOP Ativa

[Home](#)[Relatórios](#)[Fale Conosco](#)[Sobre a Plataforma](#)[Servidores Cadastrados](#)

Bem vindo a plataforma UFOP Ativa !

Lorem ipsum dolor sit amet

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Donec quis facilisis felis. Nullam sed fringilla massa. Suspendisse a dui ligula. Pellentesque augue turpis, dapibus eu turpis nec, tristique pellentesque nisi. Mauris finibus lectus sed enim luctus interdum. Vestibulum imperdiet risus at vehicula tempor.

Phasellus bibendum maximus vehicula. In urna suscipit, gravida ex vel, dictum nunc. Duis egestas tempor arcu. Nulla ac ultrices nibh. Donec consequat blandit lectus, efficitur aliquam elit ullamcorper non. Praesent quis suscipit dolor, non euismod neque. Nunc turpis nunc, volutpat nec molestie a, rutrum vel ligula.



Relatórios

[Home](#)

[Relatórios](#)

[Fale Conosco](#)

[Sobre a Plataforma](#)

[Servidores Cadastrados](#)

- ☐ Filtro 1
- ☒ Filtro 2
- ☐ Filtro 3
- ☐ Filtro 4
- ☒ Filtro 5
- ☐ Filtro 6

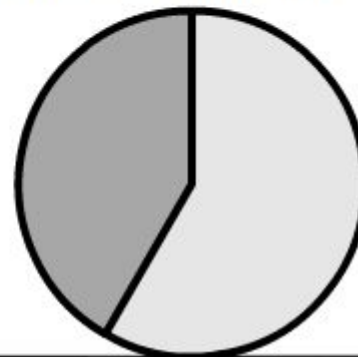
Emitir Relatório

Gerar Gráficos

Gerar Tabelas

Número de professores efetivos	900
Número de professores efetivos com Lattes	878
Números de Unidades	15
Número de Departamentos	10

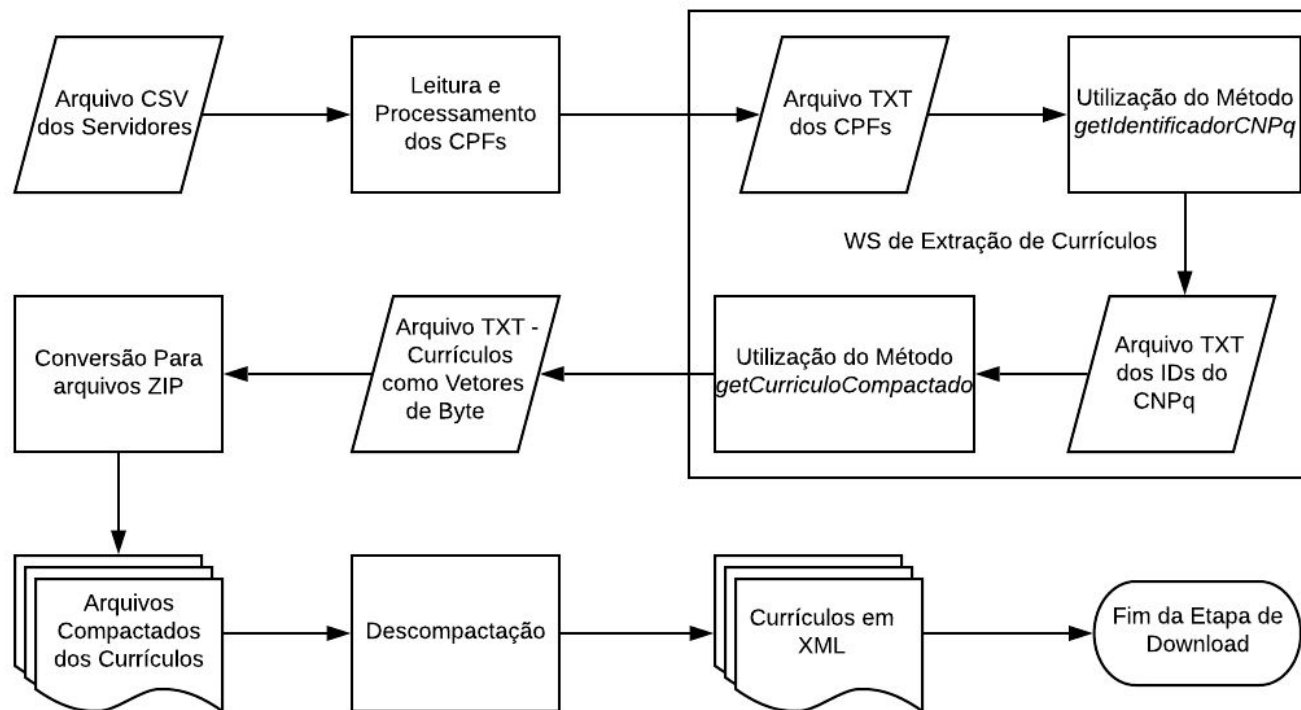
Professores por unidade



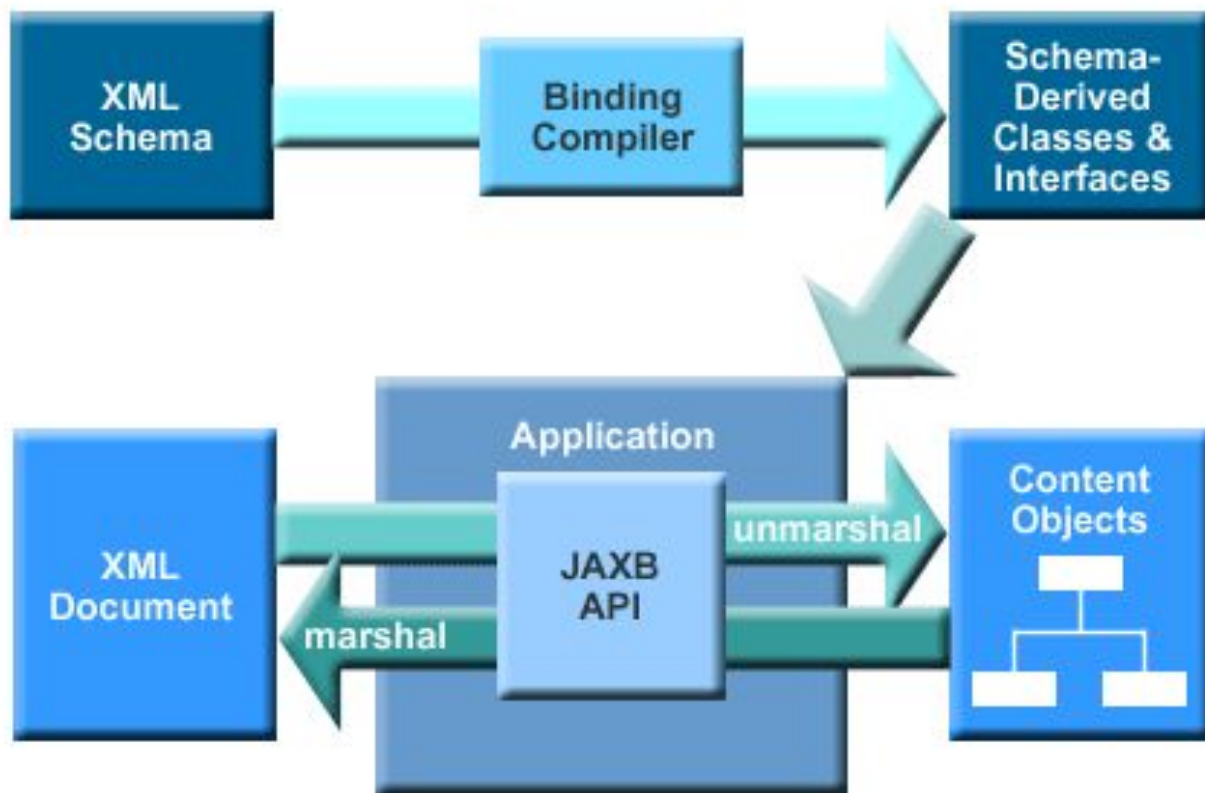
Publicações



Extração dos Dados - Download dos Currículos

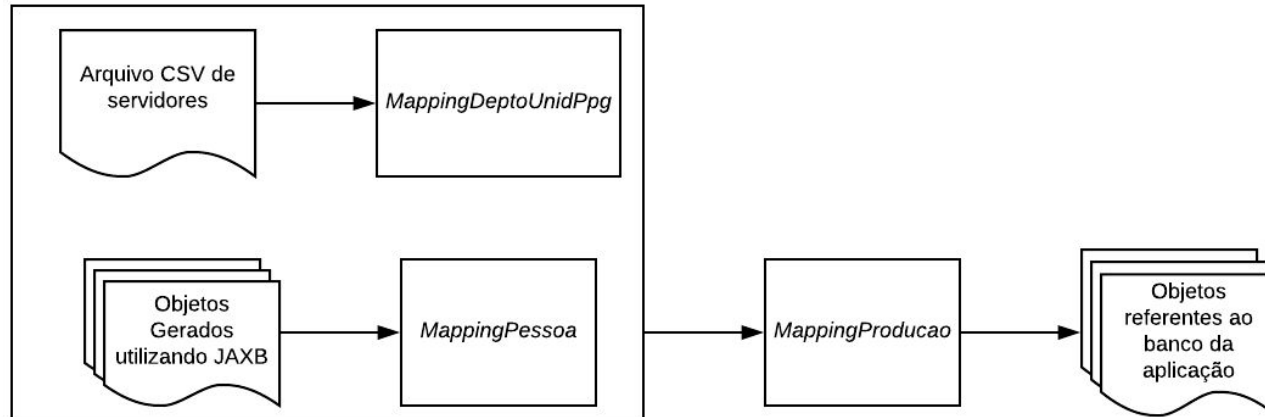


Extração dos Currículos em XML - JAXB



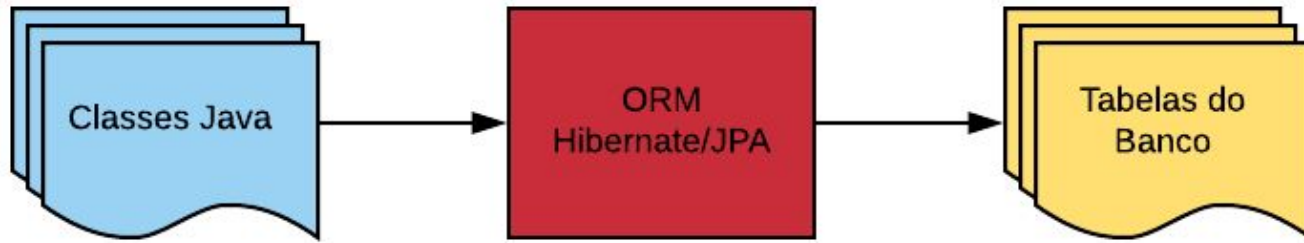
Mapeamento para Objetos Java

- Basicamente o que é feito nos objetos que representam o banco do sistema é a chamada de métodos *setters* para salvar atributos comuns ou *add* para colocar em listas de objetos. Esses dados posteriormente são mapeados e armazenados no banco de dados.



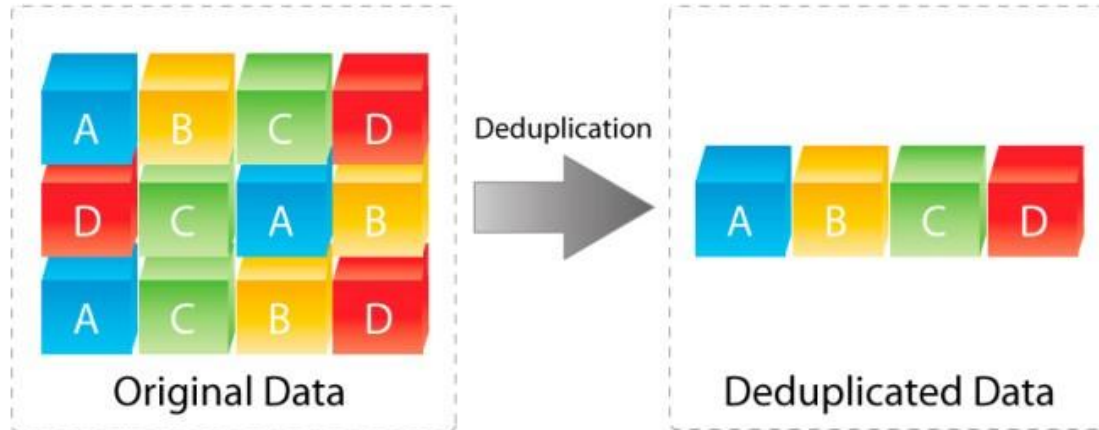
Mapeamento Objeto-Relacional (Hibernate e JPA)

- O Hibernate é o framework ORM, ou seja, a implementação física do que será utilizado para persistir, remover, atualizar ou buscar dados no SGBD;
- JPA é uma camada que descreve uma interface comum para frameworks ORM.



Deduplicação de Produções

- A deduplicação consiste basicamente em eliminar cópias duplicadas de dados com o intuito principal de reduzir a sobrecarga de armazenamento;
- Nesta primeira versão, a deduplicação foi realizada nos artigos completos publicados em periódicos.





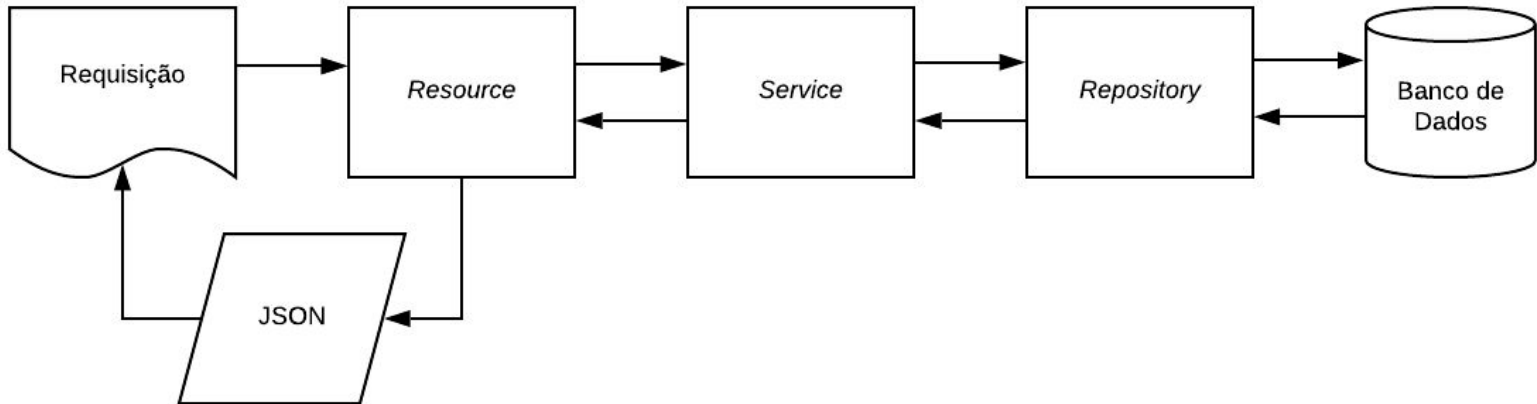
Deduplicação de Produções

- Verifica-se a similaridade de dois artigos de maneira hierárquica;
- Primeiramente, se ambos apresentarem DOI, é verificado se possuem o mesmo valor para este campo;
- Caso contrário, utiliza-se um índice de similaridade baseado na distância de edição para analisar a similaridade do título e do periódico ou revista. A similaridade é aceita se a comparação superar o limiar de 0.9 definido.

$$\textit{similaridade} = 1 - \frac{\textit{distEdicao}}{\textit{maiorTermo}}$$

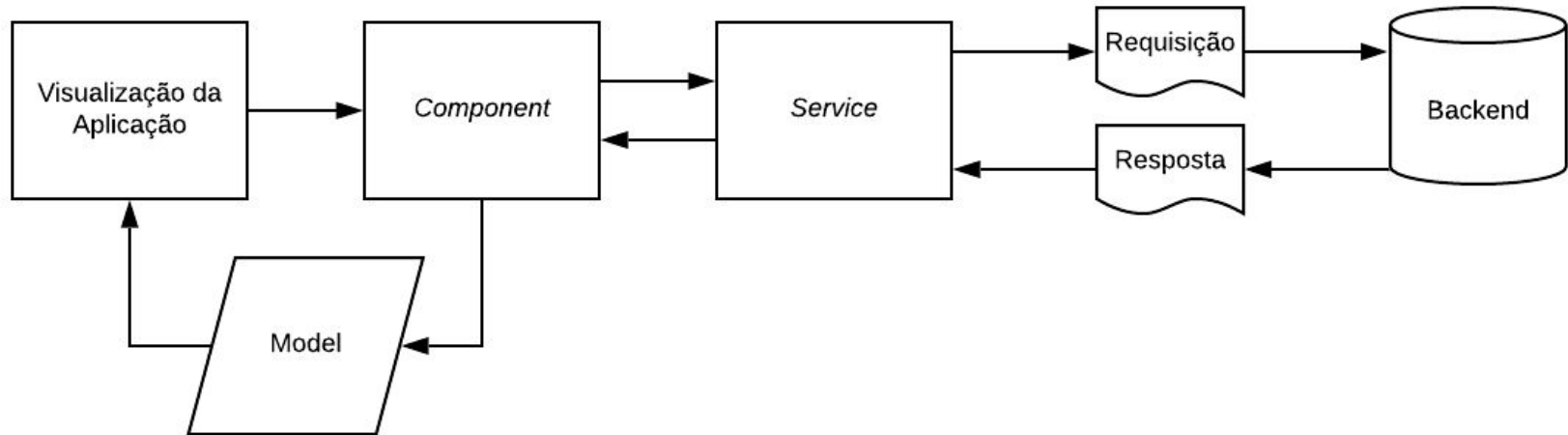
Implementação do UFOP Ativa - *Backend*

- *Config;*
- *Repositories;*
- *Resources;*
- *Services.*



Implementação do UFOP Ativa - *Frontend*

- *Components*;
- *Services*;
- *Models*.



Faça uma busca

Bem vindo à plataforma UFOP Ativa!

Aqui você pode visualizar de maneira simples e intuitiva informações relacionadas as competências da UFOP.

ESTUDO DE CASO





UFOP Ativa

Início

Dados Relevantes

Fale Conosco

Sobre o Sistema

Servidor Docente

João da Silva Vieira

UNIDADE: INSTITUTO DE CIÊNCIAS EXATAS E BIOLÓGICAS (ICES)
DEPARTAMENTO: DEPARTAMENTO DE COMPUTAÇÃO
LATTES: lattex.cpqg.br/000000000000000001
ÚLTIMA ATUALIZAÇÃO DO LATTES: 28/11/2019
NOMES DE CITAÇÕES: Vieira, J. S. / VIEIRA, João S. / VIEIRA, JOÃO SILVA

Visualizar Gráficos

Formação Acadêmica

1998	Graduação	Engenharia Mecatrônica	Universidade Federal de São João Del Rey
2002	Mestrado	Ciências da Computação	Universidade Federal de Ouro Preto
2015	Doutorado	Ciências da Computação	Universidade Federal de Ouro Preto

Orientações Concluídas

Mestrado: 7

Doutorado: 2

Pós-Doutorado: 0

Outras: 35

Projetos de Pesquisa em Andamento

2006 - Atual	Ambiente Colaborativo para Análise de Imagens
2012 - Atual	Processamento Digital de Imagens aplicados na área da saúde
2018 - Atual	Elaboração de Sistemas de Indicação

Produção Bibliográfica

Artigos Aceitos para Publicação (2)

2019

Study of Security Camera Images Processing Applied in Crime Identification

2017

CNN's Application in Biometric Recognition from Ear Imaging

Artigos Publicados (13)

Trabalhos em Eventos (2)

Outras Produções Bibliográficas (40)

Produção Técnica

Software (1)

2005

Ambiente de Melhoria de Qualidade de Imagens

Produto Tecnológico (4)

Outras Produções Técnicas (31)

Especialidades

Processamento de Imagens

Reconhecimento de Padrões

Grande Área	Área	Sub-área
Ciências Exatas e da Terra	Ciência da Computação	Metodologia e Técnicas da Computação

Linguagens de Programação

Palavras Chave

Palavra	Nº Utilizações
processamento de imagens	14
filmes ultrafinos	9
visão computacional	8
simulação computacional	6
polymers	5
dosimeter	4

Dados Relevantes

NÚMERO DE SERVIDORES DA UFOP	1653
NÚMERO DE SERVIDORES COM LATTES	1369 (82,81%)
NÚMERO DE UNIDADES CADASTRADAS	12
NÚMERO DE DEPARTAMENTOS CADASTRADOS	137

Unidades

#	Unidade
01	INSTITUTO DE CIENCIAS EXATAS E BIOLOGICAS
02	ESCOLA DE MINAS
03	INSTITUTO DE CIENCIAS HUMANAS E SOCIAIS
04	INSTITUTO DE CIENCIAS EXATAS E APLICADAS
05	INSTITUTO DE CIENCIAS SOCIAIS E APLICADAS
06	ESCOLA DE MEDICINA
07	ESCOLA DE DIREITO, TURISMO E MUSEOLOGIA
08	INSTITUTO DE FILOSOFIA ARTES E CULTURA
09	ESCOLA DE FARMACIA
10	ESCOLA DE NUTRICAO
11	CENTRO DE EDUCACAO ABERTA E A DISTANCIA
12	CENTRO DESPORTIVO

Unidade: INSTITUTO DE CIENCIAS EXATAS E BIOLOGICAS (ICEB)

Departamentos: 7

Servidores: 273

Departamentos

DEPARTAMENTO DE QUIMICA

DEPARTAMENTO DE MATEMATICA

DEPARTAMENTO DE CIENCIAS BIOLOGICAS

DEPARTAMENTO DE FISICA

DEPARTAMENTO DE COPMUTACAO

DEPARTAMENTO DE BIODIVERSIDADE, EVOLUCAO E MEIO AMBIENTE

DEPARTAMENTO DE ESTATISTICA

Departamento: DEPARTAMENTO DE COMPUTACAO

Unidade: **INSTITUTO DE CIENCIAS EXATAS E BIOLOGICAS**

Servidores (66) Por ordem alfabética

1	ALAN ROBERT RESENDE DE FREITAS
2	ALEXANDRE MAGNO DE SOUSA
3	ALVARO RODRIGUES PEREIRA JUNIOR
4	AMANDA SAVIO NASCIMENTO E SILVA
5	ANDERSON ALMEIDA FERREIRA
6	ANDREA GOMES CAMPOS BIANCHI
7	BRUNO CERQUEIRA HOTT
8	BRUNO PEREIRA DOS SANTOS
9	BRUNO RABELLO MONTEIRO
10	CARLOS ALBERTO MARQUES PIETROBON
11	CARLOS FREDERICO MARCELO DA CUNHA CAVALCANTI
12	CARLOS HENRIQUE GOMES FERREIRA
13	CRISTIANO AMARO DA MATTA
14	DARLAN NUNES DE BRITO
15	DAYANNE GOUVEIA COELHO
16	DIEGO ZUQUIM GUIMARAES
17	EDUARDO DA SILVA RIBEIRO
18	EDUARDO JOSE DA SILVA LUZ
19	ELTON JOSE DA SILVA

CONCLUSÃO E TRABALHOS FUTUROS





Conclusão

- Neste trabalho desenvolveu-se uma versão inicial do UFOP Ativa, um sistema que auxilia a exibição e a análise de dados acadêmicos de servidores da UFOP, através de informações provenientes da Plataforma Lattes;
- O sistema denominado permite que os usuários tenham acesso a informações acerca de pesquisas de servidores, unidades e departamentos da universidade, em um ambiente amigável e de fácil entendimento.



Conclusão

- É notável que a base de dados de currículos do Lattes é uma das bases de pesquisadores mais completas do mundo e compreende uma vasta quantidade de informações diversificadas;
- Essas informações podem ser exploradas e manipuladas para o levantamento de diversos estudos e dados estatísticos relevantes, para as instituições de ensino em geral.



Conclusão

- A elaboração deste projeto é de grande valia para a UFOP e pode ser evoluído através de melhorias e adições;
- O software apresentado é apenas o começo de uma ferramenta que é extremamente significativa para a instituição.



Trabalhos Futuros

- Perspectiva de usuário administrador e suas funcionalidades;
- Extração de informações automática periodicamente (ou manualmente quando se julgar necessário);
- Mapeamento de coautorias;
- Mapeamento completo das especialidades;
- Elaboração de testes unitários e de integração para aplicação;
- Segurança do sistema.

Perguntas

