



# Sistema Nou-Rau

## Arquivamento Indexação de Documentos Digitais



Rubens Queiroz de Almeida  
queiroz@ccuec.unicamp.br





# Motivação

- Criação de um sistema para facilitar o armazenamento e indexação de documentos digitais
  - Moldado para participação de voluntários através da Internet
  - Criação de um repositório de documentos sobre software livre
  - Disseminar conhecimento





# Nou-Rau é Software Livre

- Uma longa história, mas
  - Criação de uma ferramenta útil para a comunidade de ensino
  - Colaboração de voluntários para seu desenvolvimento
  - Quanto mais usuários ...
    - Produto cada vez melhor 
    - Todos ganham: a universidade, escolas, a comunidade, o país, você





# Software Livre é Liberdade

- Liberdade de executar o programa, para qualquer fim, da forma que se desejar
- Liberdade para modificar o programa e adaptá-lo às suas necessidades
- Liberdade para ajudar seu vizinho, distribuindo cópias do programa
- Liberdade para ajudar a criar sua comunidade, publicando versões aperfeiçoadas, permitindo que outros se beneficiem de seu trabalho





# Software Livre

é

Compartilhamento  
Amizade  
Solidariedade  
Um mundo Melhor!





**COPIE!!!**



**Seja Livre**

**Use Software Livre**

# Personagens do Sistema



**Administrador**



**Responsável**



**Colaborador**

**Sistema**

---



**Visitante**



# Princípios Básicos

- Armazenamento de qualquer tipo de documento
- Armazenar informações sobre o documento
- Oferecer recursos de busca avançados
- Controlar a aceitação de novos documentos
- Simples de se usar e manter





# Quem é Quem?

- Visitante
  - Navega, busca e faz download de documentos
- Colaborador
  - Faz a submissão de documentos
- Responsável
  - Aprova documentos
- Administrador
  - Gerente geral do sistema. Cria tópicos, aprova novos colaboradores, etc.





# Responsável

- Aprova a inclusão de um novo documento
- Cada tópico tem apenas um responsável
- Um responsável pode cuidar de vários tópicos





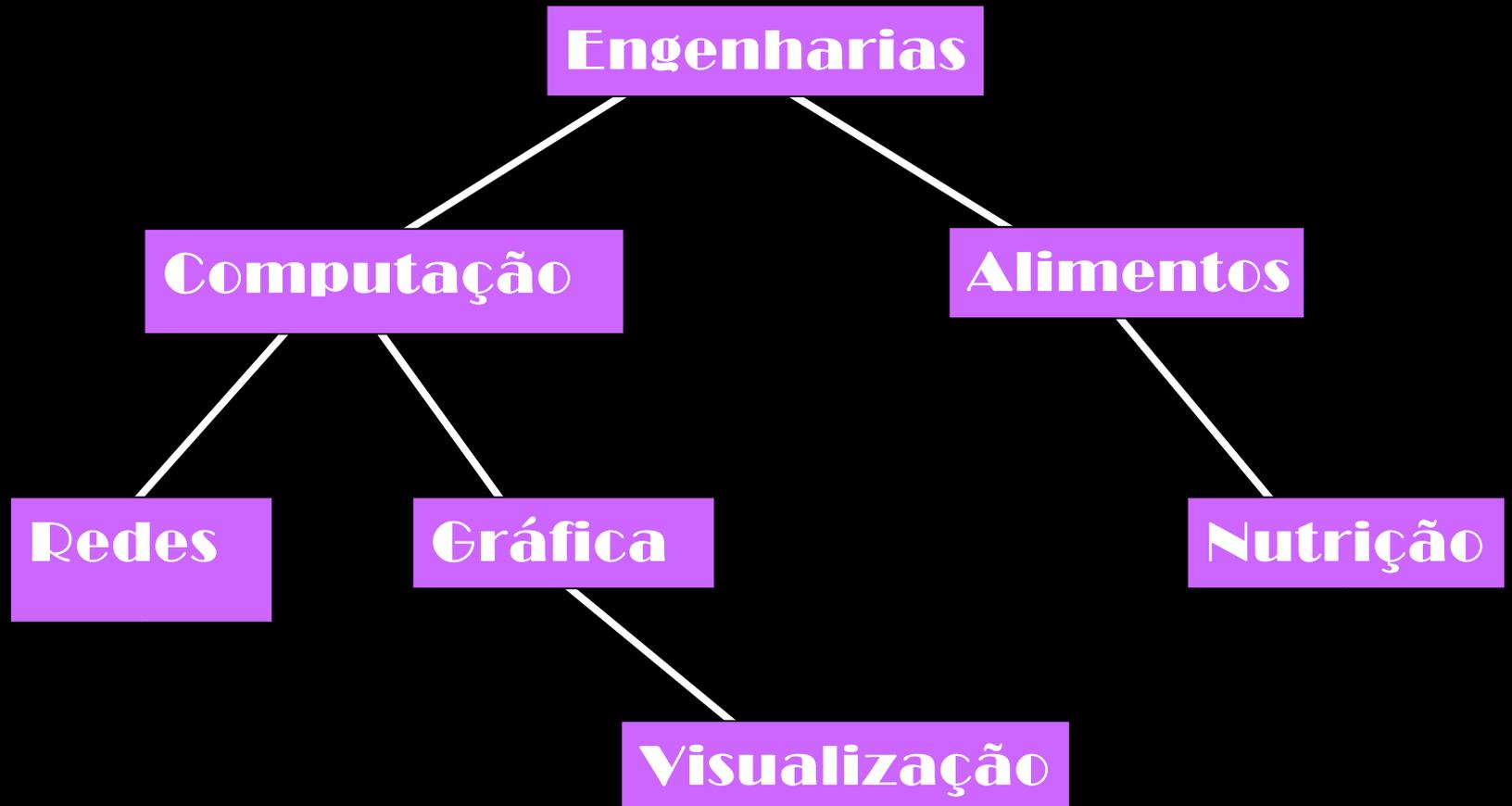
# Administrador

- Faz a manutenção do sistema
- Cria novos tópicos
- Define categorias e formatos





# Estrutura de Tópicos





# Documentos



- Título, autor(es), e-mail(s), palavras-chave, descrição e versão
- Informações adicionais 
- Dono, categoria e formato 
- Arquivo, tamanho, data de criação, data de alteração e número de downloads





# Controle de Publicação



- É preciso estar registrado como colaborador para enviar um documento
- O documento deve ter um formato compatível com uma das categorias do tópico
- O documento deve ser aprovado pelo responsável





# Categorias e Formatos

- Um tópico pode aceitar diversos tipos de documentos:
  - Teses, artigo, figuras, tabelas, entrevistas, filmes, músicas, ...
- Uma categoria define um ou mais formatos válidos
  - PDF, DOC, GIF, JPEG, MPEG, MP3, ...





# Mecanismo de Busca

- Busca realizada sobre as informações sobre o documento e sobre seu texto completo
- Suporte para qualquer formato para o qual exista um conversor
  - ASCII, PDF, DOC, SGML, WML, LaTeX, TeX, PostScript, Microsoft Word, PowerPoint, Excel, OpenOffice, ...





# Outros Recursos

- Compressão interna 
- Possibilidade de bloquear documentos
- Mecanismo de busca ajustável
- Novos formatos podem ser entendidos





# Possíveis usos

- Biblioteca digital
  - Unicamp, Unesp, MEC, ...
- Organização de documentos pessoais
- Intranets
  - Banco de Currículos
  - Documentos empresariais
  - Artigos técnicos
  - Páginas pessoais
- Muito mais ....





# Tecnologia Utilizada

Sistemas  
Abertos

Ferramentas  
livres e gratuitas

Estabilidade,  
qualidade

Liberdade



Antiword



xpdf

pdftotext



The Apache Software Foundation

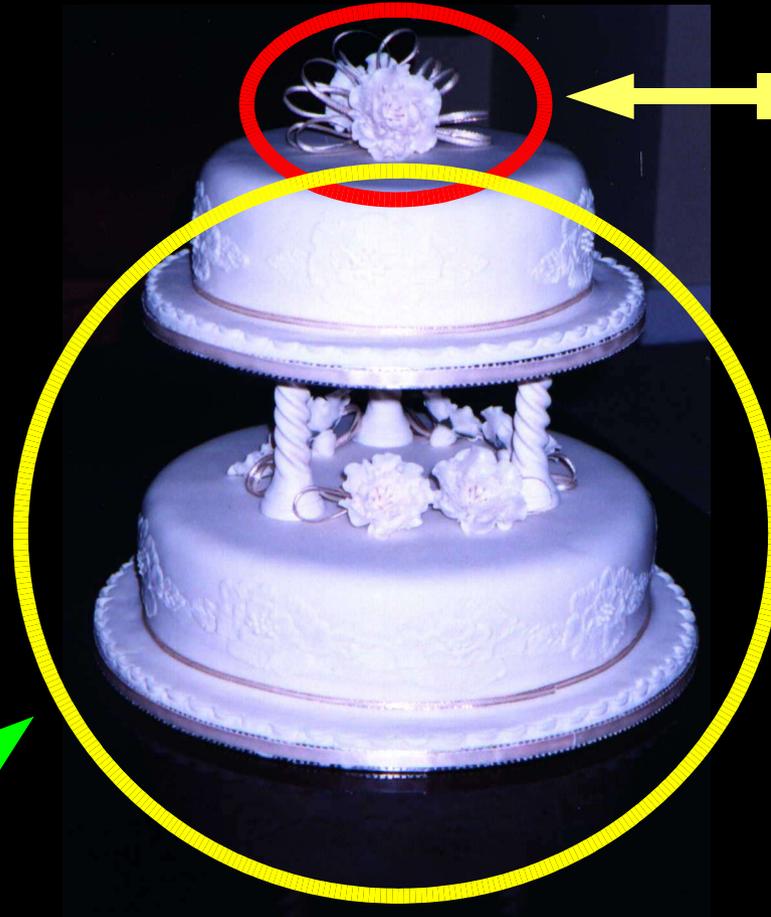
<http://www.apache.org/>



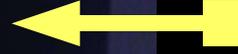


# Software Livre e Inovação

Software  
Livre



Nou-Rau





# Software Livre e Inovação

<http://www.dicas-l.unicamp.br/dicas-l/20040718.php>

| APLICATIVO     | LINHAS       | CARACTERES    | %           |
|----------------|--------------|---------------|-------------|
| Antiword       | 44491        | 1444534       | 0,87        |
| Apache         | 40053        | 4165130       | 2,51        |
| dvi2tty        | 7245         | 226850        | 0,14        |
| htdig          | 461943       | 11648195      | 7,02        |
| perl           | 1530939      | 50113497      | 30,2        |
| php            | 836506       | 24436796      | 14,73       |
| Postgresql     | 1787926      | 61370089      | 36,99       |
| pstotext       | 4734         | 144623        | 0,09        |
| recode         | 205215       | 8000820       | 4,82        |
| xlhtml         | 37830        | 1167848       | 0,7         |
| xpdf           | 101163       | 2846137       | 1,72        |
| <b>nou-rau</b> | <b>11203</b> | <b>354583</b> | <b>0,21</b> |





# Agradecimentos

Instituto Vale do Futuro  
Centro de Computação da  
Unicamp

Biblioteca Central da Unicamp





# Contato

**Rubens Queiroz de Almeida**  
queiroz@ccuec.unicamp.br

**Dicas-L**

<http://www.Dicas-L.unicamp.br>

**Esta Apresentação**

<http://www.Dicas-L.unicamp.br/cursos>

