Integração de Dados

Disciplina de Procedência de Dados e
Data Warehousing

Profa. Dra. Cristina Dutra de Aguiar Ciferri
cdac@icmc.usp.br

Tópicos

- Trabalho desenvolvido
  - ferramenta Reconciliador v.1
  - ferramenta Reconciliador v.2
  - Sistema Urano

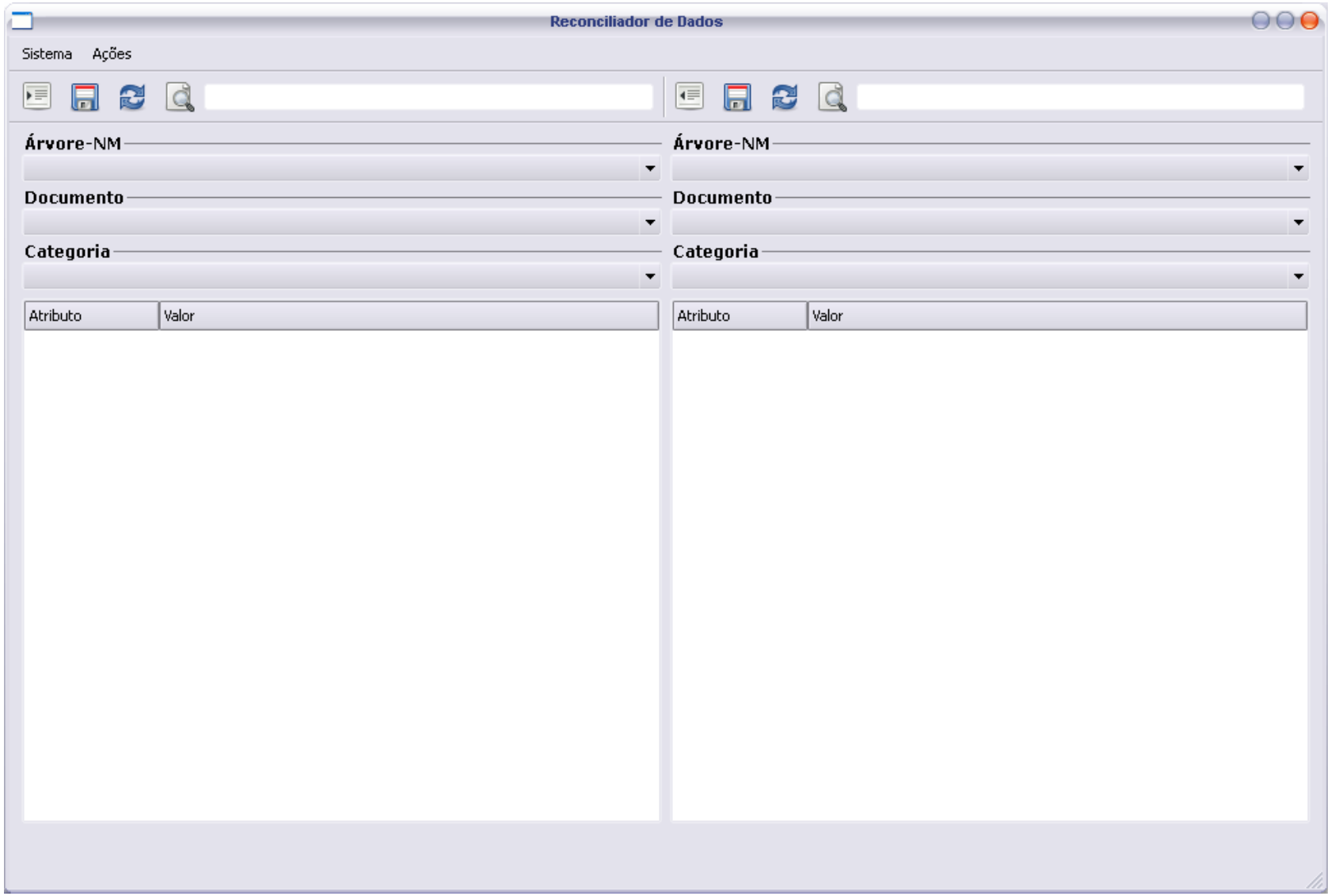
Ferramenta Reconciliador v.1

- Voltada à integração de instâncias (conflitos de valores)
- Características
  - semi-automatiza a identificação de
    - objetos correspondentes
    - inconsistências
  - ajuda o usuário a
    - eliminar inconsistências
    - completar dados
    - trocar dados

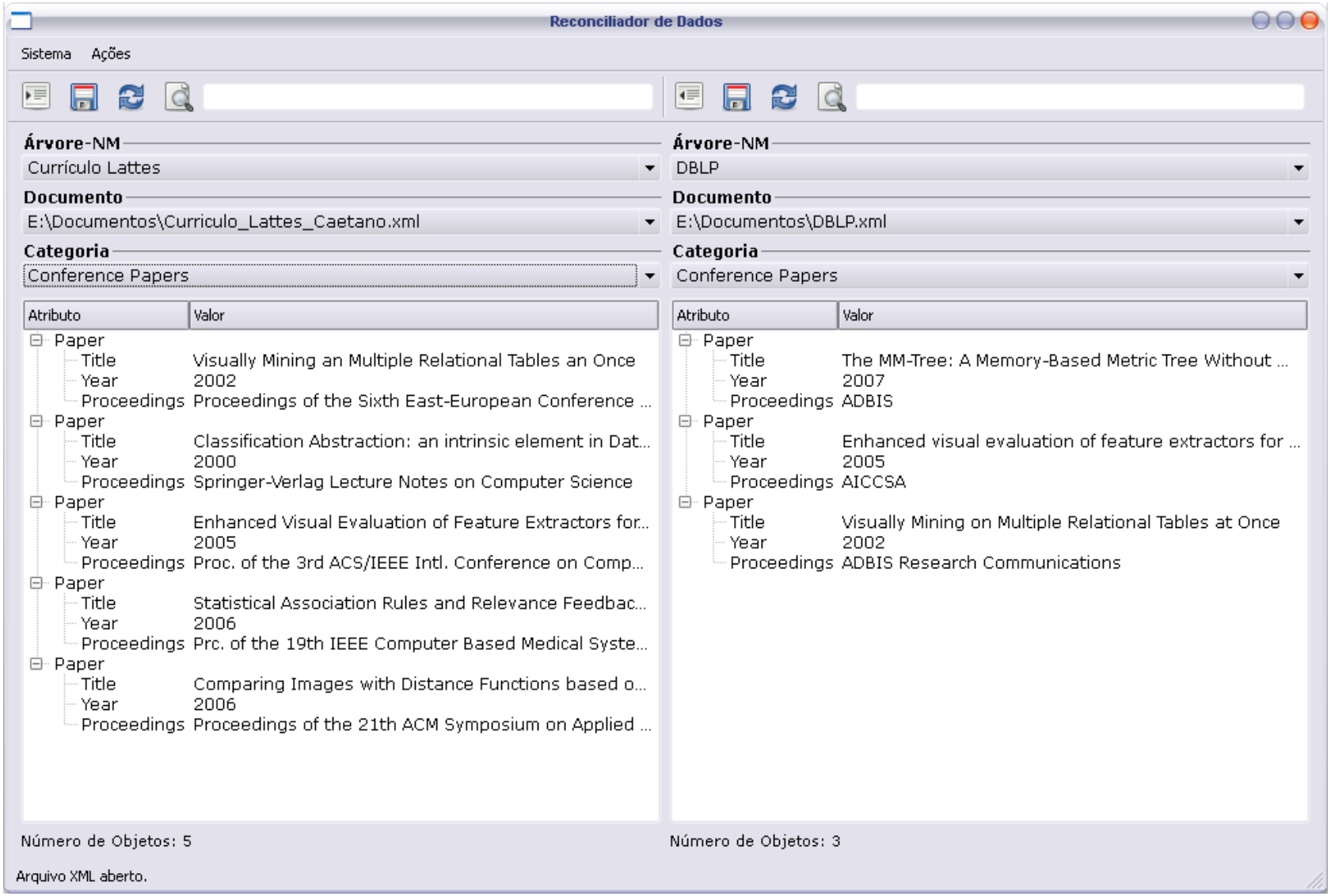
Ferramenta Reconciliador v.1

- Aspectos funcionais
  - visualização dos objetos de dois documentos
    - visualização lado-a-lado
  - sincronização de objetos
  - edição de dados incompletos ou errados
  - troca de dados entre objetos de diferentes documentos

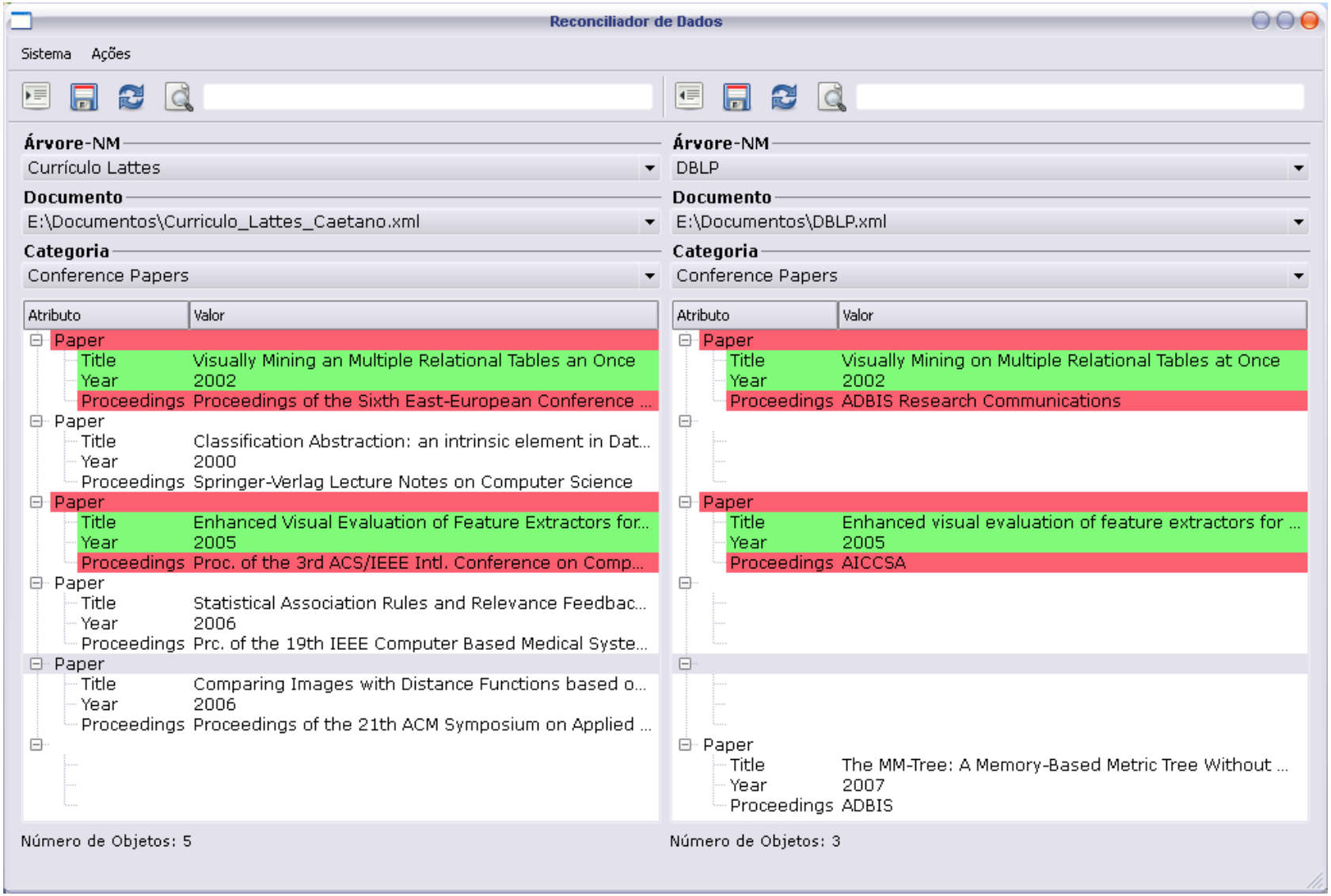
Interface


Reconciliador de Dados
Sistema
Ações
Árvore-NM
Documento
Categoria
Atributo
Valor
Árvore-NM
Documento
Categoria
Atributo
Valor

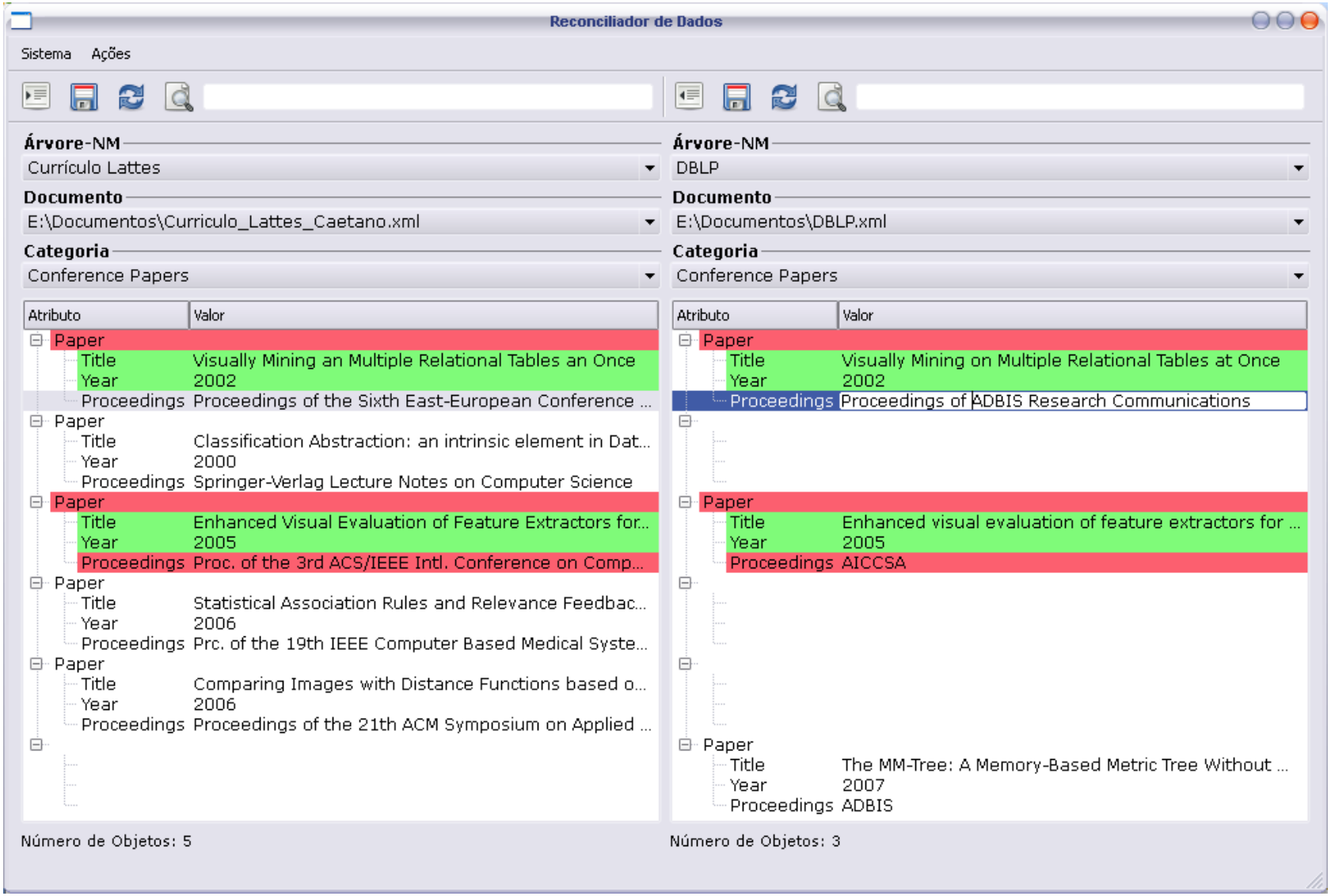
Visualização



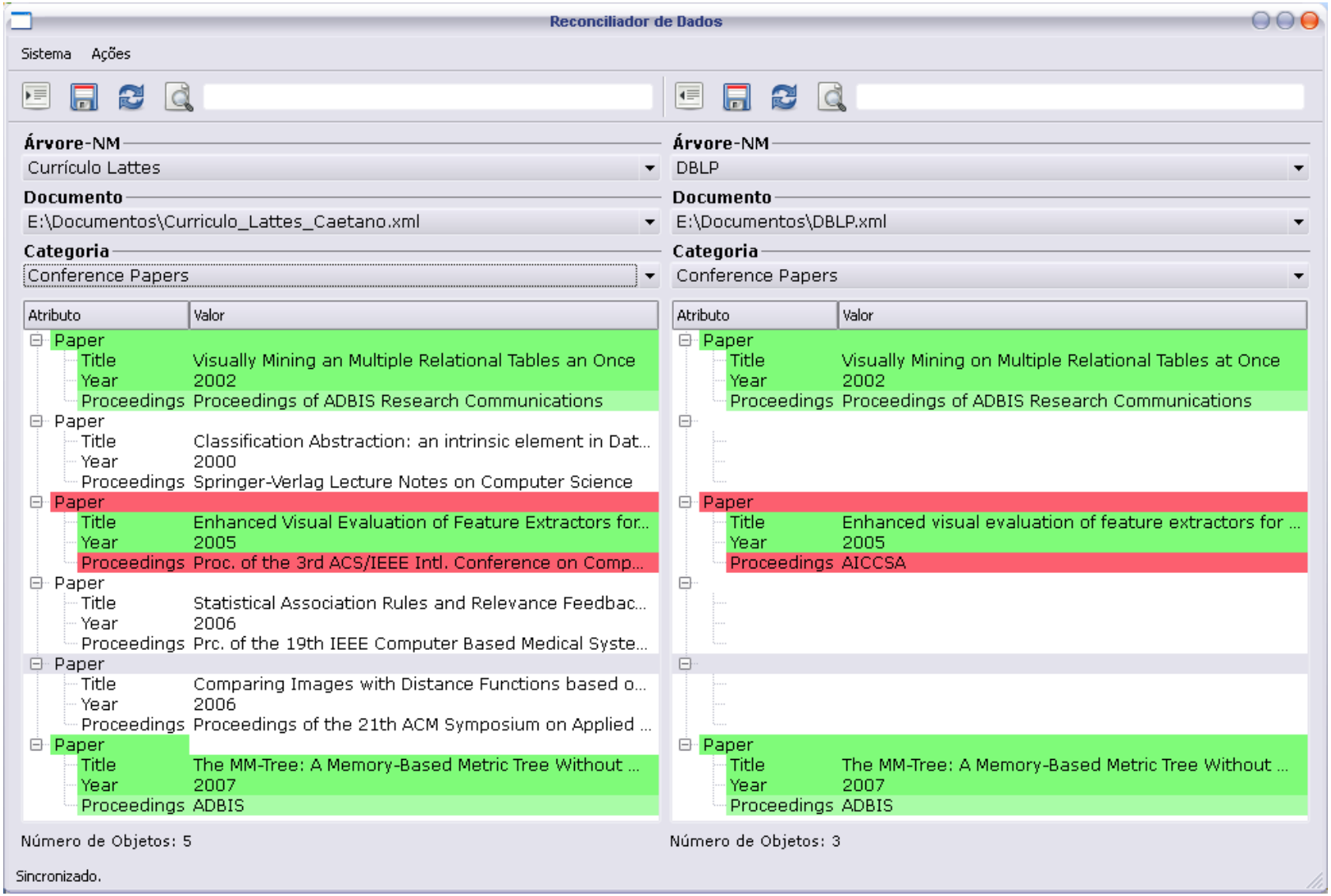
Sincronização



Edição


Reconciliador de Dados
Sistema
Ações
Árvore-NM
Currículo Lattes
Documento
E:\Documentos\Curriculo_Lattes_Caetano.xml
Categoria
Conference Papers
Atributo
Valor
Paper
Title Visually Mining an Multiple Relational Tables an Once
Year 2002
Proceedings Proceedings of the Sixth East-European Conference ...
Paper
Title Classification Abstraction: an intrinsic element in Dat...
Year 2000
Proceedings Springer-Verlag Lecture Notes on Computer Science
Paper
Title Enhanced Visual Evaluation of Feature Extractors for...
Year 2005
Proceedings Proc. of the 3rd ACS/IEEE Intl. Conference on Comp...
Paper
Title Statistical Association Rules and Relevance Feedbac...
Year 2006
Proceedings Prc. of the 19th IEEE Computer Based Medical Syste...
Paper
Title Comparing Images with Distance Functions based o...
Year 2006
Proceedings Proceedings of the 21th ACM Symposium on Applied ...
Número de Objetos: 5
Árvore-NM
DBLP
Documento
E:\Documentos\DBLP.xml
Categoria
Conference Papers
Atributo
Valor
Paper
Title Visually Mining on Multiple Relational Tables at Once
Year 2002
Proceedings Proceedings of ADBIS Research Communications
Paper
Title Enhanced visual evaluation of feature extractors for ...
Year 2005
Proceedings AICCSA
Paper
Title The MM-Tree: A Memory-Based Metric Tree Without ...
Year 2007
Proceedings ADBIS
Número de Objetos: 3

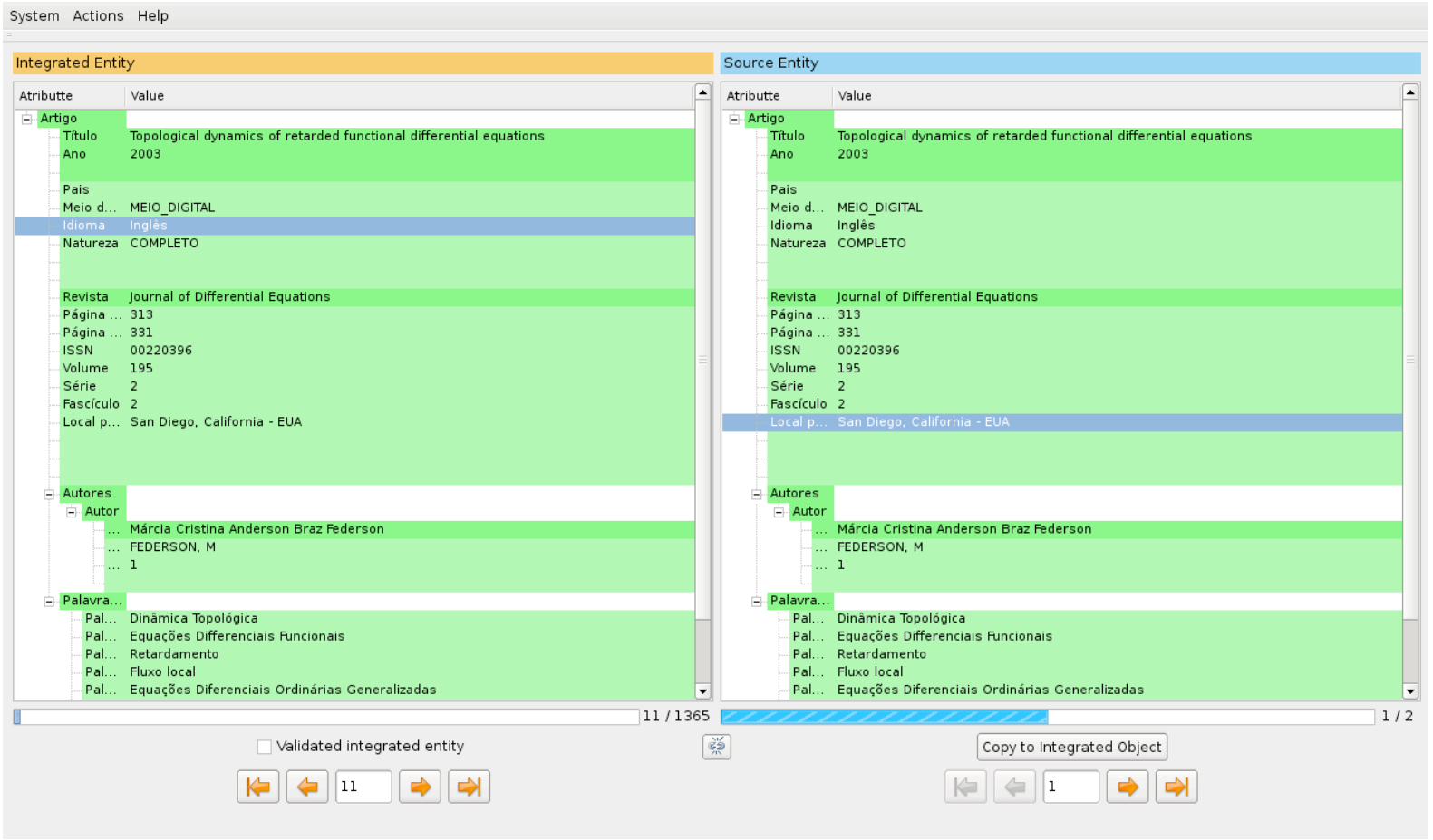
Troca de Dados



Ferramenta Reconciliador v.2

- Extensão da ferramenta Reconciliador v.1
- Voltada à integração de instâncias (resolução de entidades)
- Aspectos funcionais
  - geração automática de uma entidade integrada que melhor representa cada agrupamento
  - mudança de uma entidade de um agrupamento para outro agrupamento

Ferramenta Reconciliador v.2



entidade integrada

entidades similares que pertencem ao mesmo agrupamento

Geração de Entidades Integradas

- Primeira entidade-fonte
  - entidade escolhida para ser a entidade integrada é a primeira entidade-fonte pertencente ao agrupamento
- Entidade-fonte com maior peso
  - entidade escolhida para ser a entidade integrada é aquela que tem mais atributos preenchidos de acordo com pesos associados aos seus atributos
  - pesos são determinados de acordo com a importância do atributo para a entidade

Geração de Entidades Integradas

- Entidade-fonte de maior frequência
  - entidade escolhida para ser a entidade integrada é aquela que aparece mais vezes no agrupamento
  - são considerados apenas alguns atributos da entidade para se fazer uma comparação exata entre os valores desses atributos

Resultados

- Testes sobre currículos de docentes
  - entidade-fonte com maior peso produziu os resultados mais satisfatórios, desde que escolhe a entidade-fonte com maior número de atributos preenchidos
  - maioria dos agrupamentos possui apenas uma entidade-fonte (mesmos resultados para todas as propostas)
  - resultados de entidade-fonte de maior frequência foram praticamente os mesmos que os resultados de primeira entidade-fonte do agrupamento

Mudança de Agrupamento

- Existe outro agrupamento similar à entidade-fonte
  - o usuário pode decidir mover a entidade-fonte para o agrupamento identificado pela ferramenta ou criar um novo agrupamento com a entidade-fonte em questão
- Não é encontrado outro agrupamento similar à entidade-fonte
  - o usuário pode escolher gerar um novo agrupamento para a entidade-fonte

O Sistema Urano

- Dados acadêmicos dos pesquisadores
  - distribuídos em uma variedade de provedores
- Provedores heterogêneos
  - podem prover dados inconsistentes, redundantes ou complementares
- Relatórios administrativos
  - usualmente gerados manualmente
  - contêm basicamente a mesma informação, porém em formatos distintos

Objetivos

- Coletar dados acadêmicos de provedores de informação distintos e integrá-los em um banco de dados centralizado
- Permitir que relatórios consistentes sejam gerados automaticamente

Ajuda a tomada de decisão e a administração institucional

Provedores de Dados

- Sistemas corporativos da Universidade
  - responsável pelos dados transacionais relacionados às atividades de graduação e pós-graduação
  - exemplos: Sistema Júpiter e Sistema Fênix (http://www.sistemas.usp.br)
- Fontes externas
  - provedores públicos contendo dados acadêmicos
  - exemplos: Lattes (http://lattes.cnpq.br), DBLP

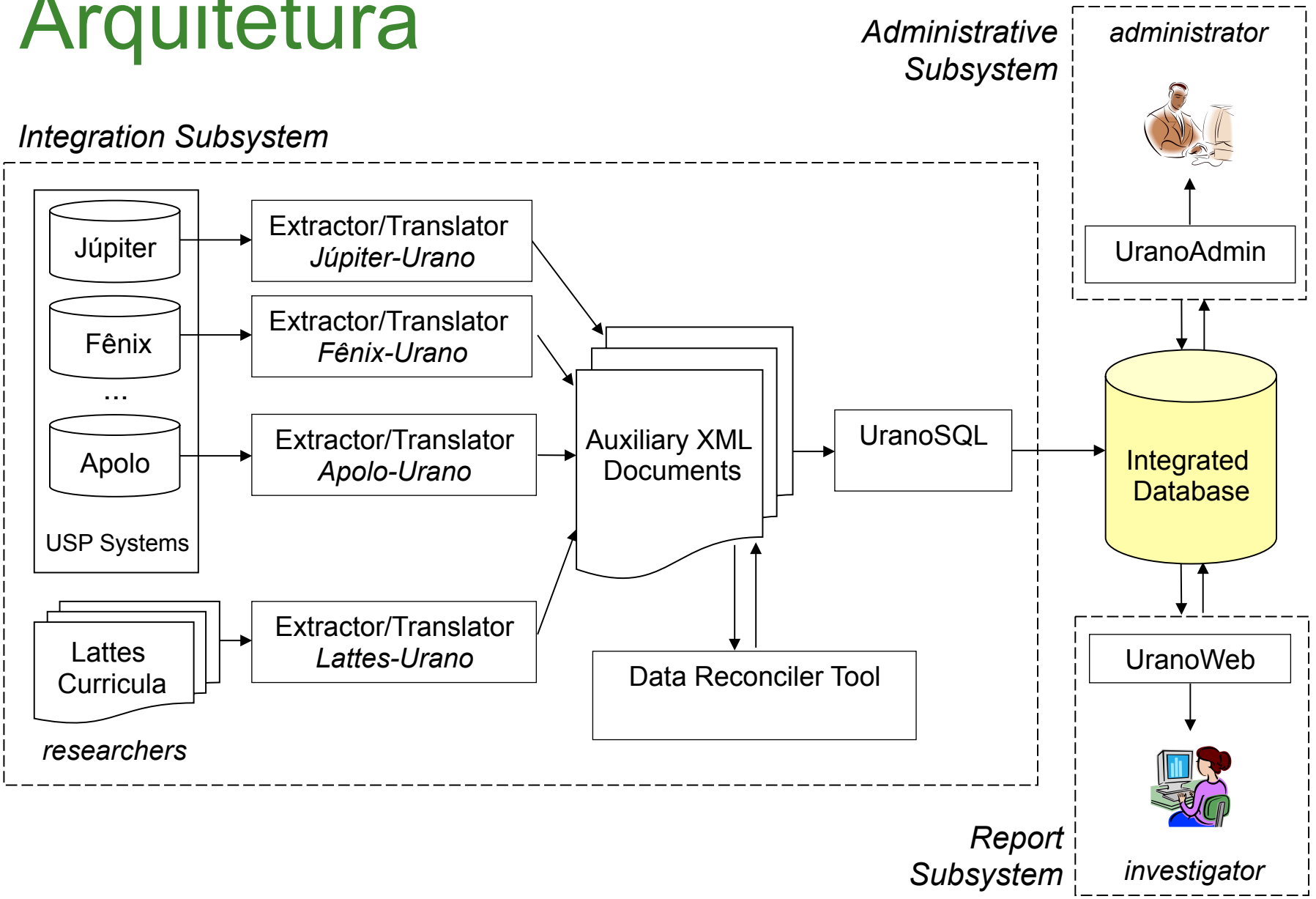
Principais Tipos de Relatório

- Docentes, seus departamentos e informações profissionais
- Publicações
  - revistas, eventos, livros, capítulos de livros
- Orientações
  - iniciação científica, mestrado, doutorado, pós-doutorado
- Participação em bancas julgadoras
  - graduação, mestrado, doutorado, concurso público

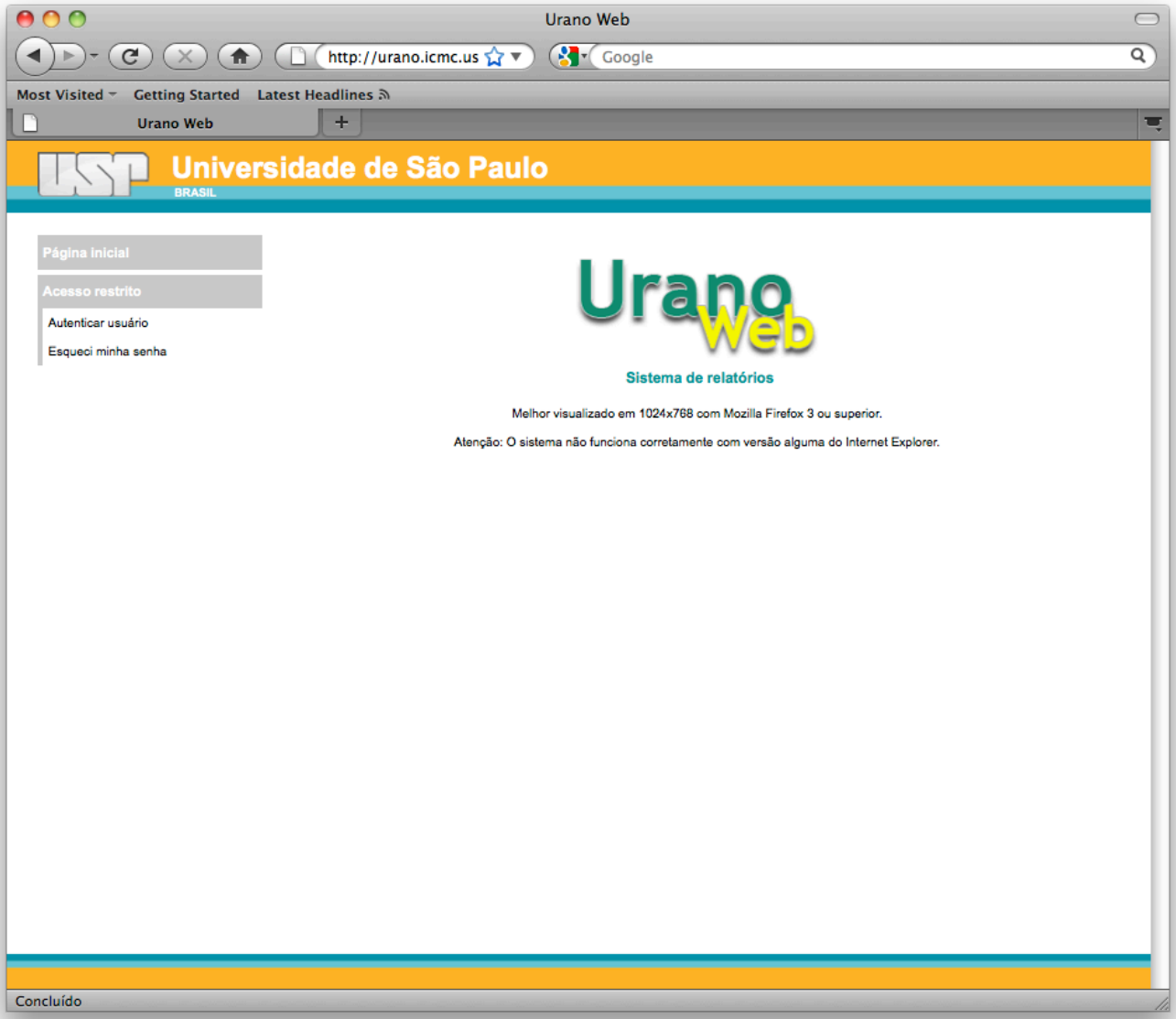
Principais Tipos de Relatório

- Participação em conferências
- Disciplinas
  - graduação e pós-graduação
- Projetos
  - descrição, órgãos financiadores

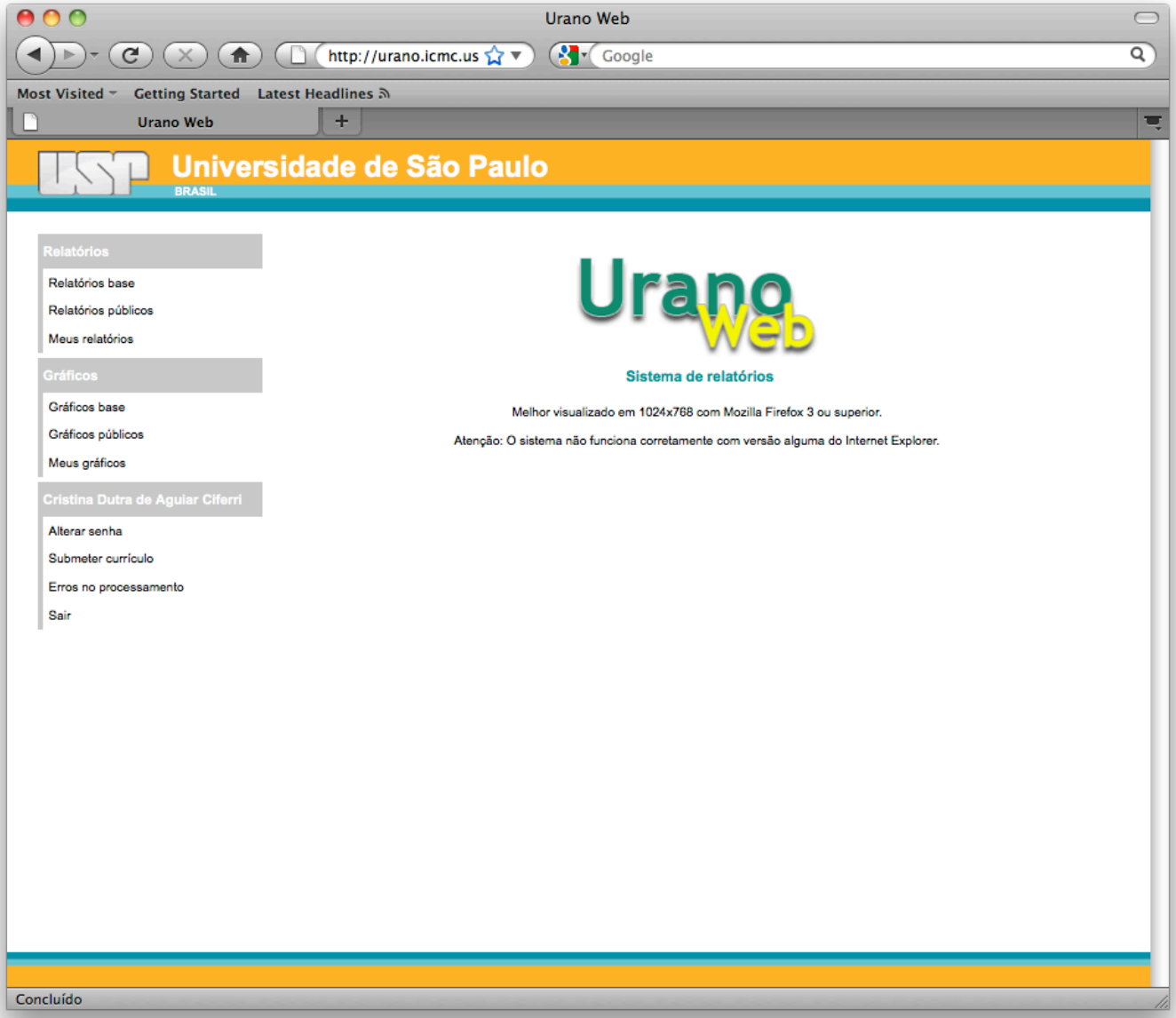
Arquitetura



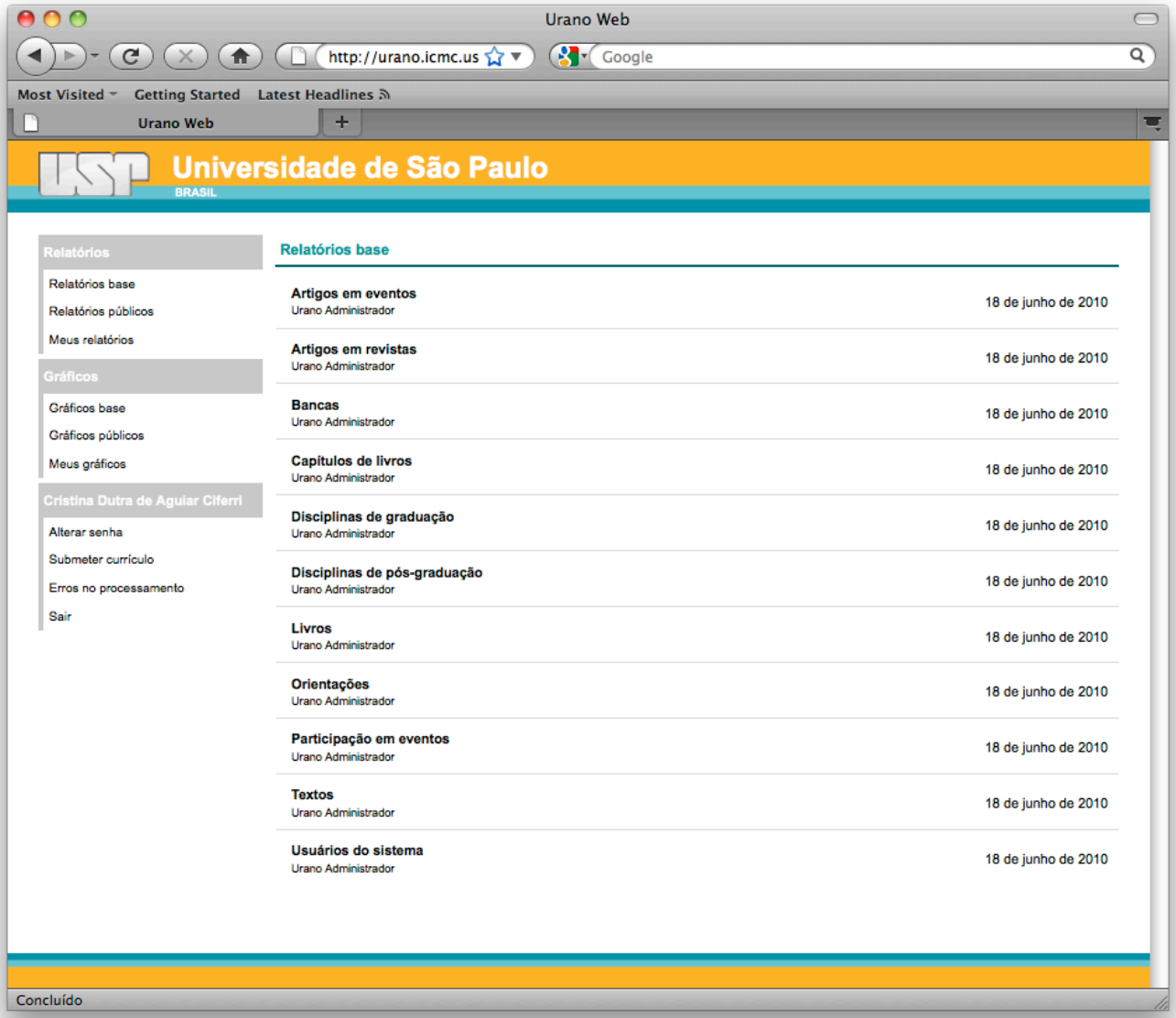
http://urano.icmc.usp.br



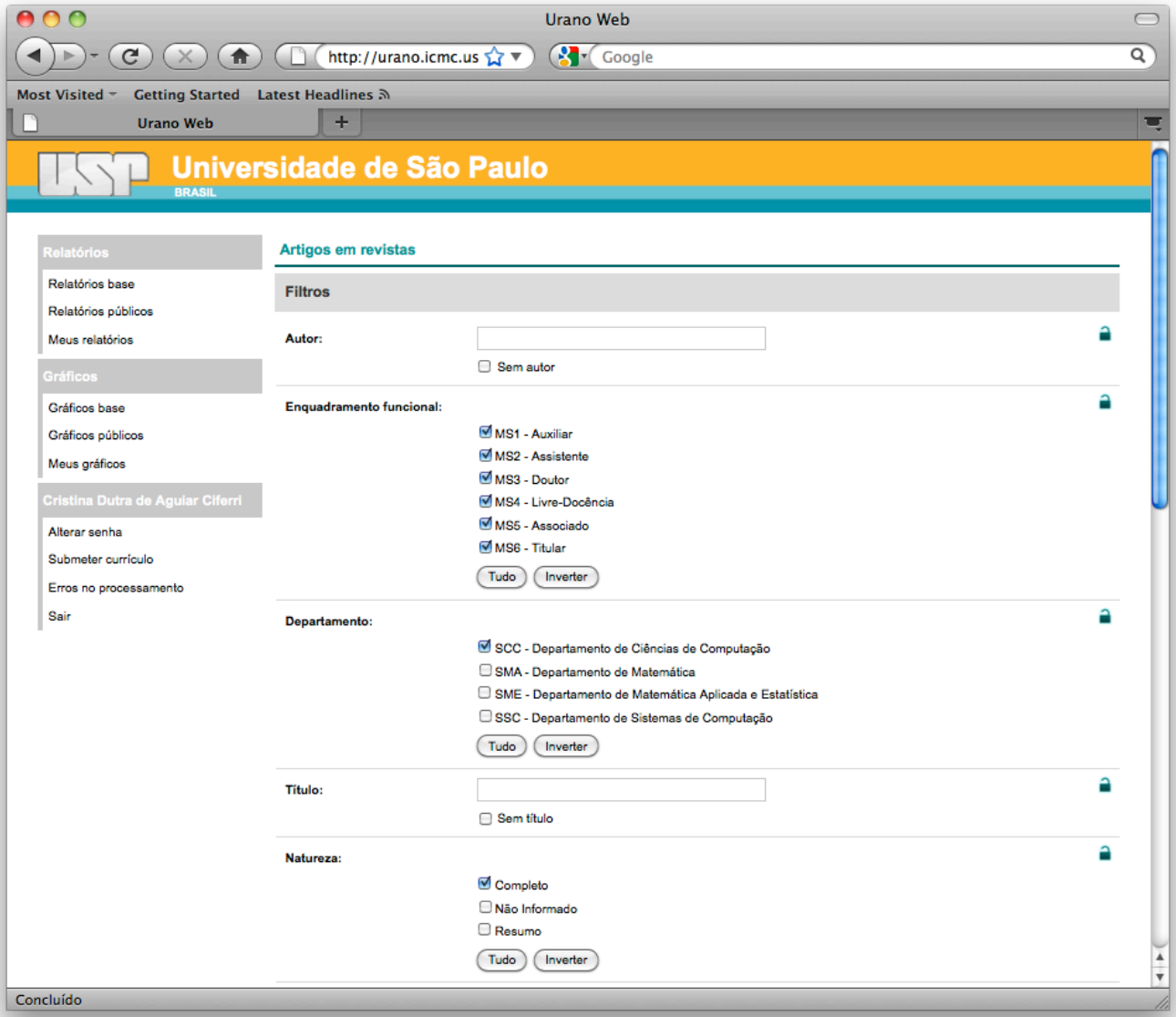
Interface


Urano Web
http://urano.icmc.us
Google
Most Visited
Getting Started
Latest Headlines
Urano Web
Universidade de São Paulo
BRASIL
Relatórios
Relatórios base
Relatórios públicos
Meus relatórios
Gráficos
Gráficos base
Gráficos públicos
Meus gráficos
Cristina Dutra de Aguiar Ciferri
Alterar senha
Submeter currículo
Erros no processamento
Sair
Urano Web
Sistema de relatórios
Melhor visualizado em 1024x768 com Mozilla Firefox 3 ou superior.
Atenção: O sistema não funciona corretamente com versão alguma do Internet Explorer.
Concluído

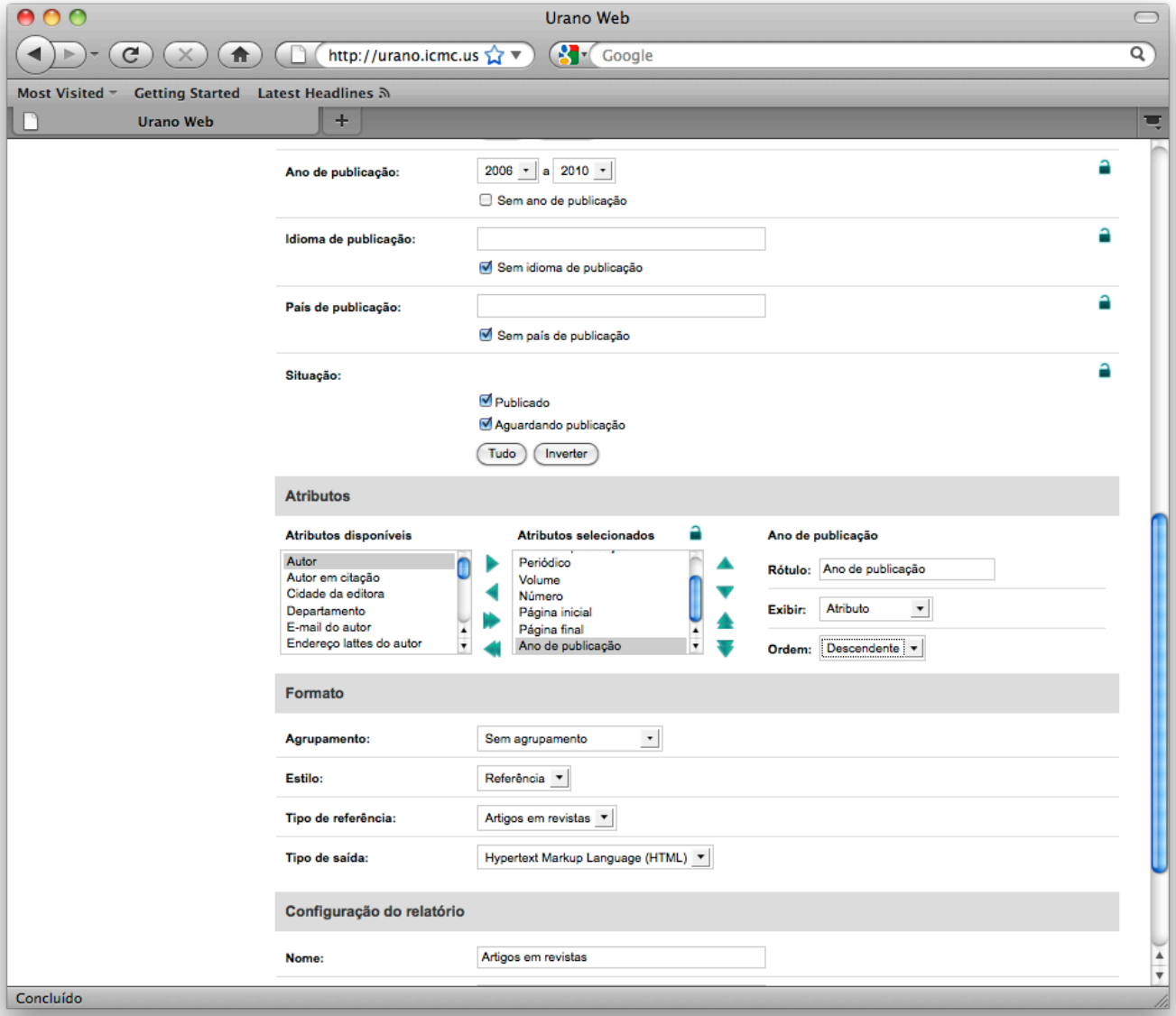
Relatórios


Urano Web
http://urano.icmc.us
Google
Most Visited
Getting Started
Latest Headlines
Urano Web
Universidade de São Paulo
BRASIL
Relatórios
Relatórios base
Relatórios públicos
Meus relatórios
Gráficos
Gráficos base
Gráficos públicos
Meus gráficos
Cristina Dutra de Aguiar Ciferri
Alterar senha
Submeter currículo
Erros no processamento
Sair
Relatórios base
Artigos em eventos
Urano Administrador
18 de junho de 2010
Artigos em revistas
Urano Administrador
18 de junho de 2010
Bancas
Urano Administrador
18 de junho de 2010
Capítulos de livros
Urano Administrador
18 de junho de 2010
Disciplinas de graduação
Urano Administrador
18 de junho de 2010
Disciplinas de pós-graduação
Urano Administrador
18 de junho de 2010
Livros
Urano Administrador
18 de junho de 2010
Orientações
Urano Administrador
18 de junho de 2010
Participação em eventos
Urano Administrador
18 de junho de 2010
Textos
Urano Administrador
18 de junho de 2010
Usuários do sistema
Urano Administrador
18 de junho de 2010
Concluído

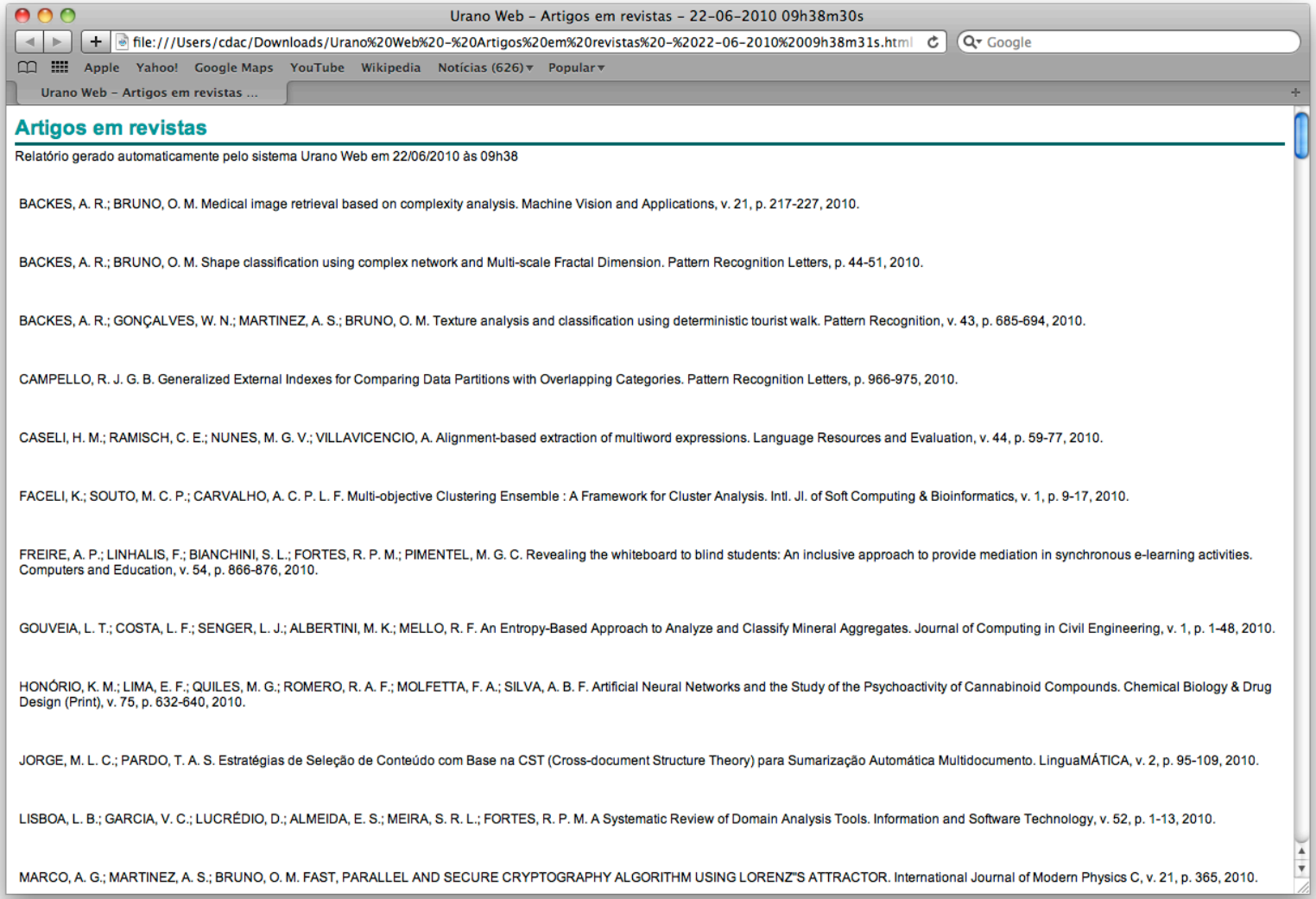
Exemplo de um Relatório



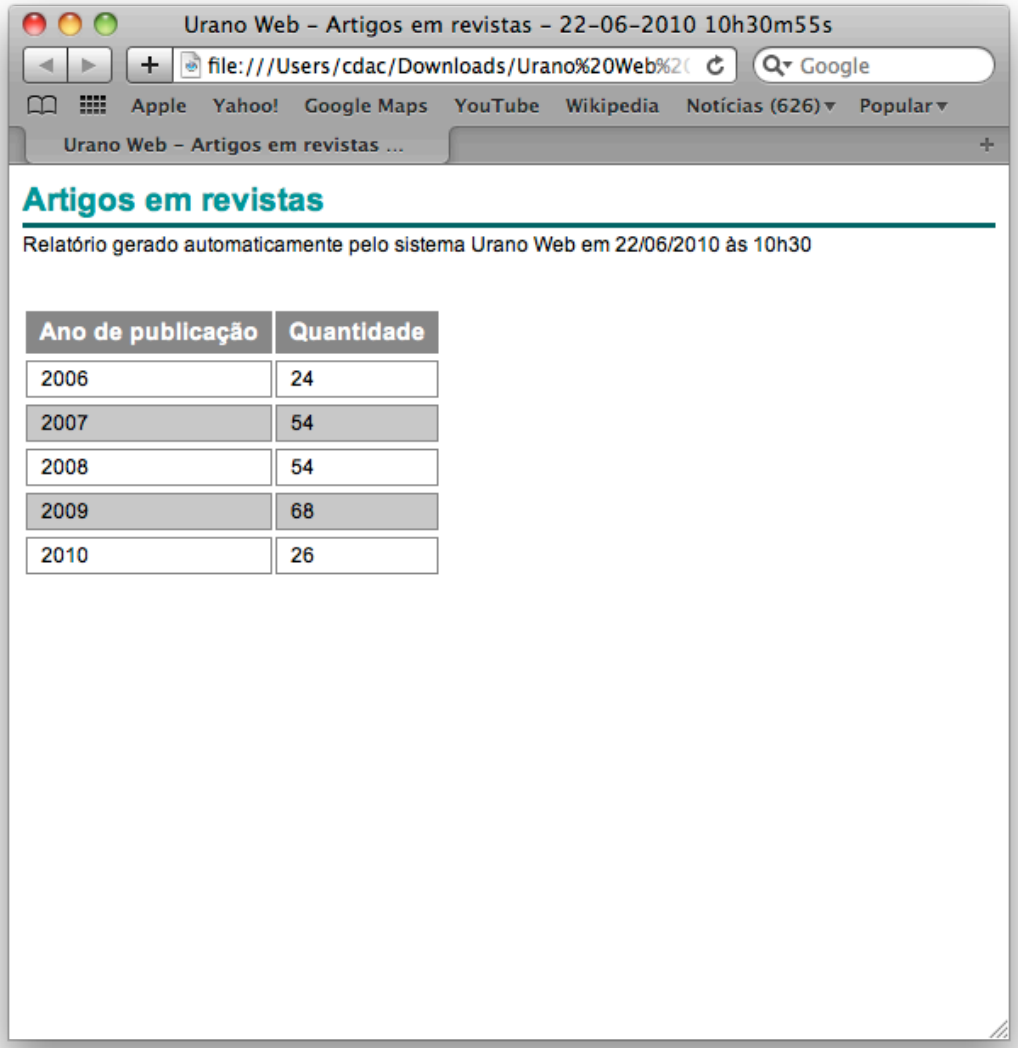
Exemplo de um Relatório



Lista de Publicações em Revista

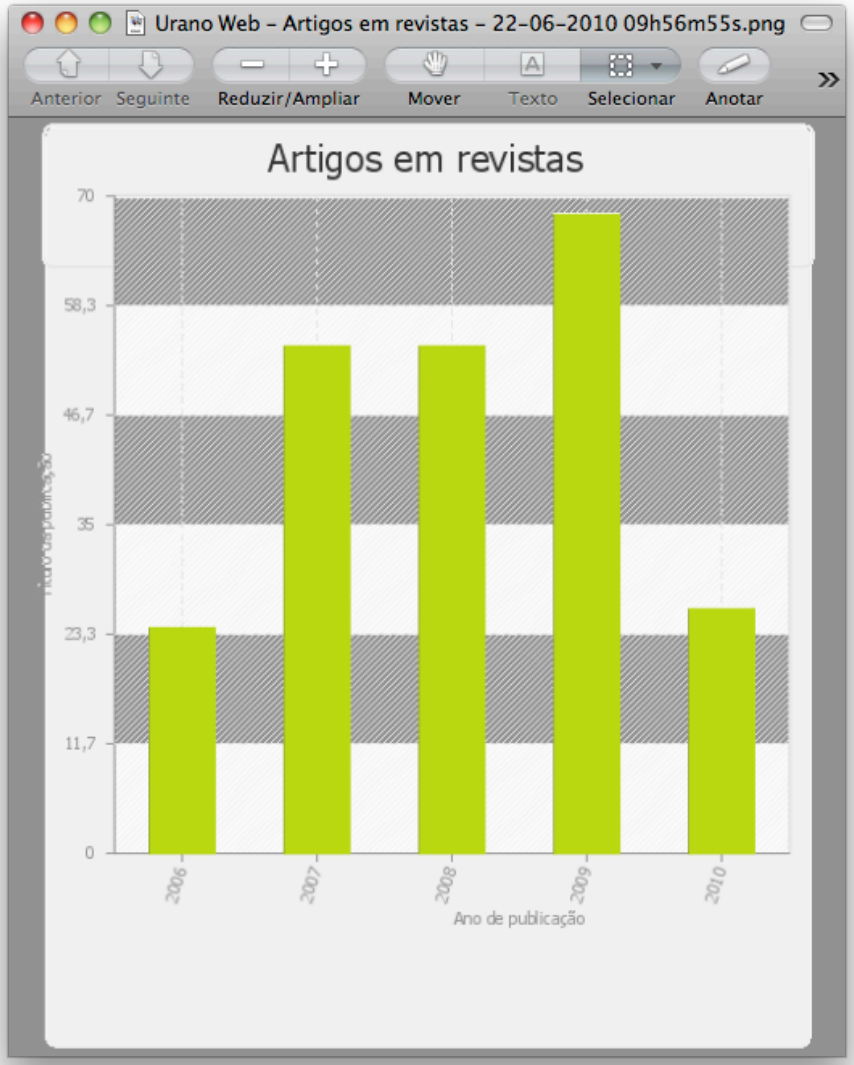


Lista de Publicações em Revista



Ano de publicação	Quantidade
2006	24
2007	54
2008	54
2009	68
2010	26

Lista de Publicações em Revista

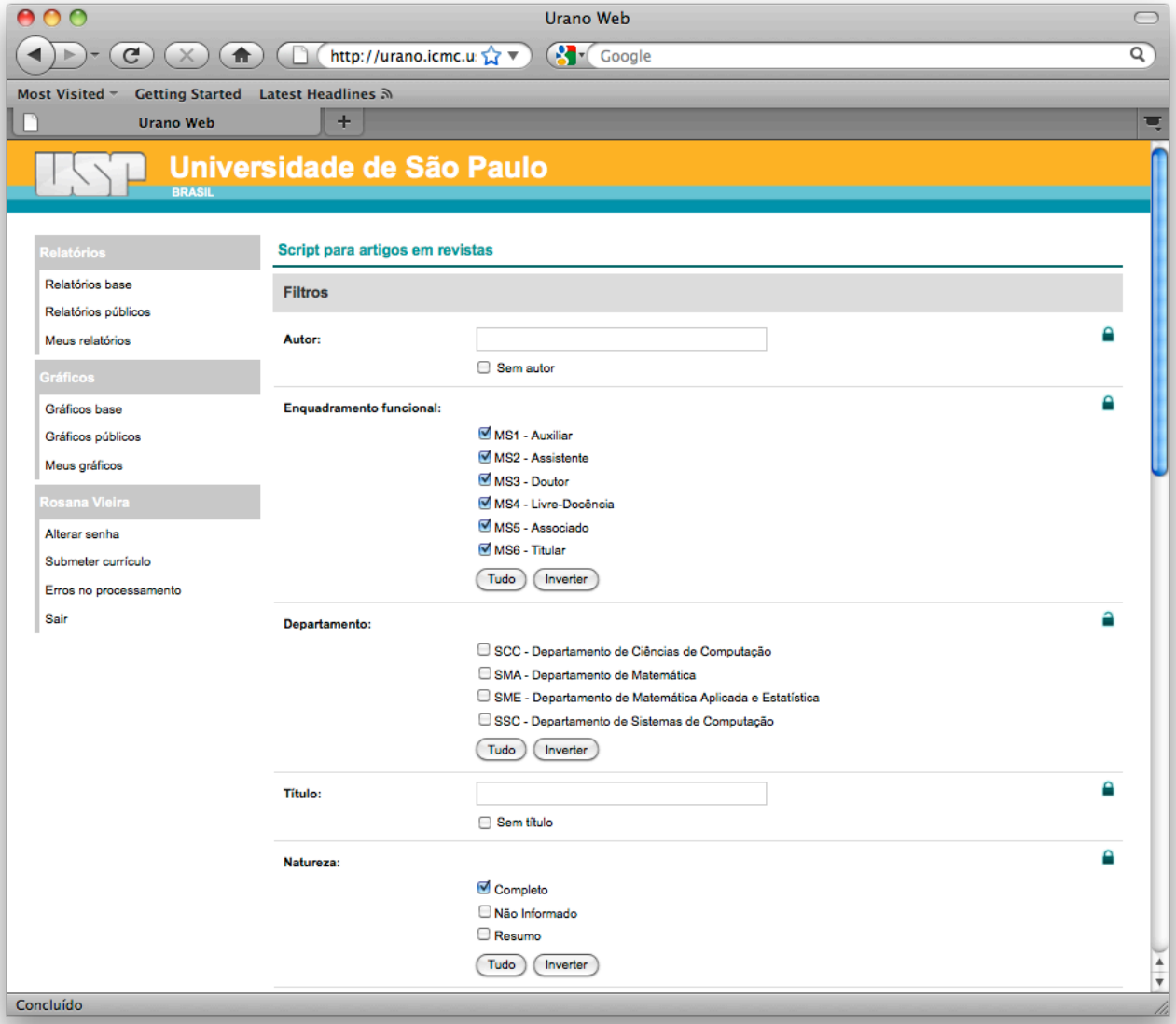


Composição de Relatórios

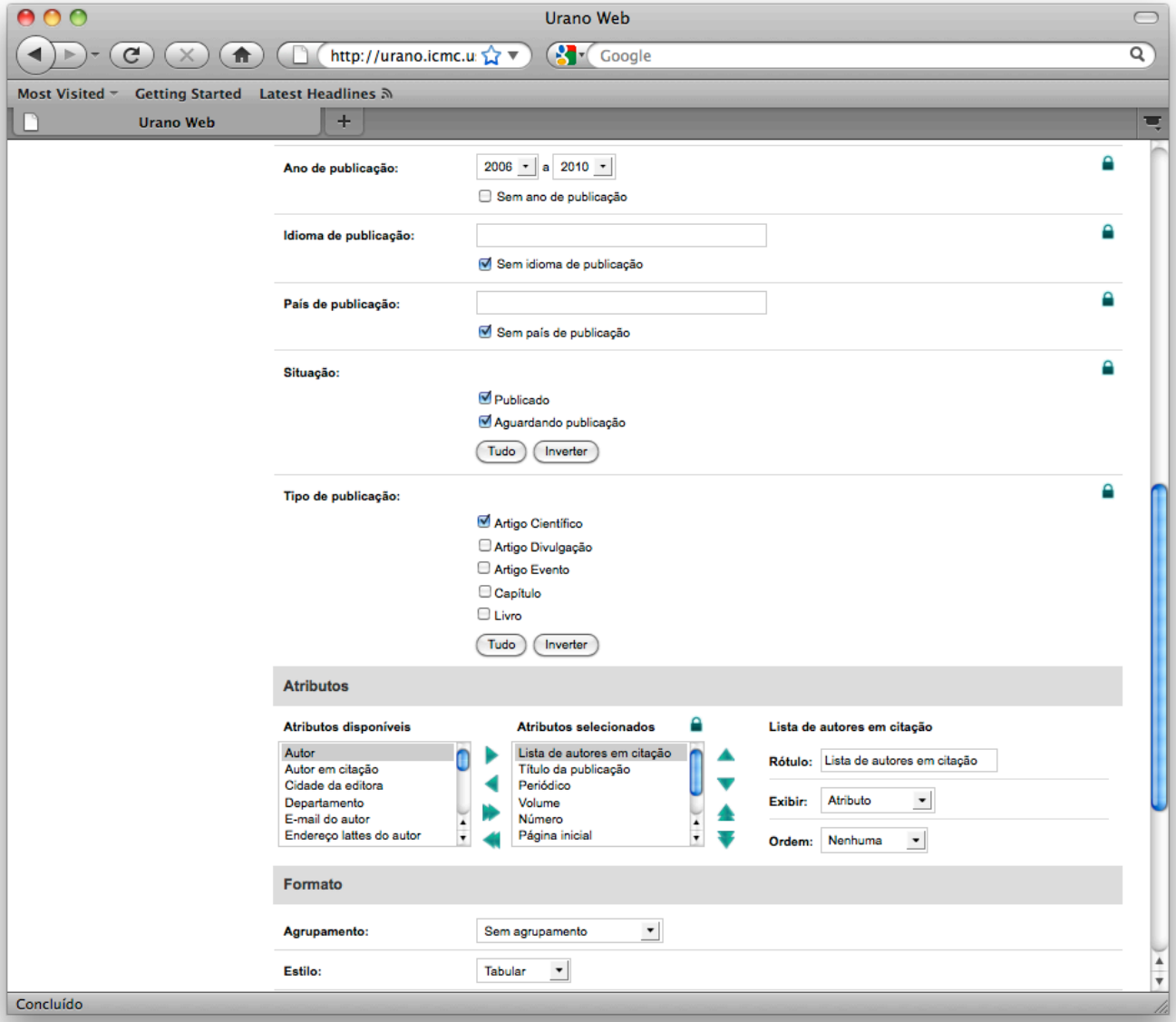
- Investigadores
  - geram o formato dos relatórios que os diferentes pesquisadores devem preencher

- Pesquisadores
  - preenchem o relatório de acordo com os seus dados acadêmicos particulares

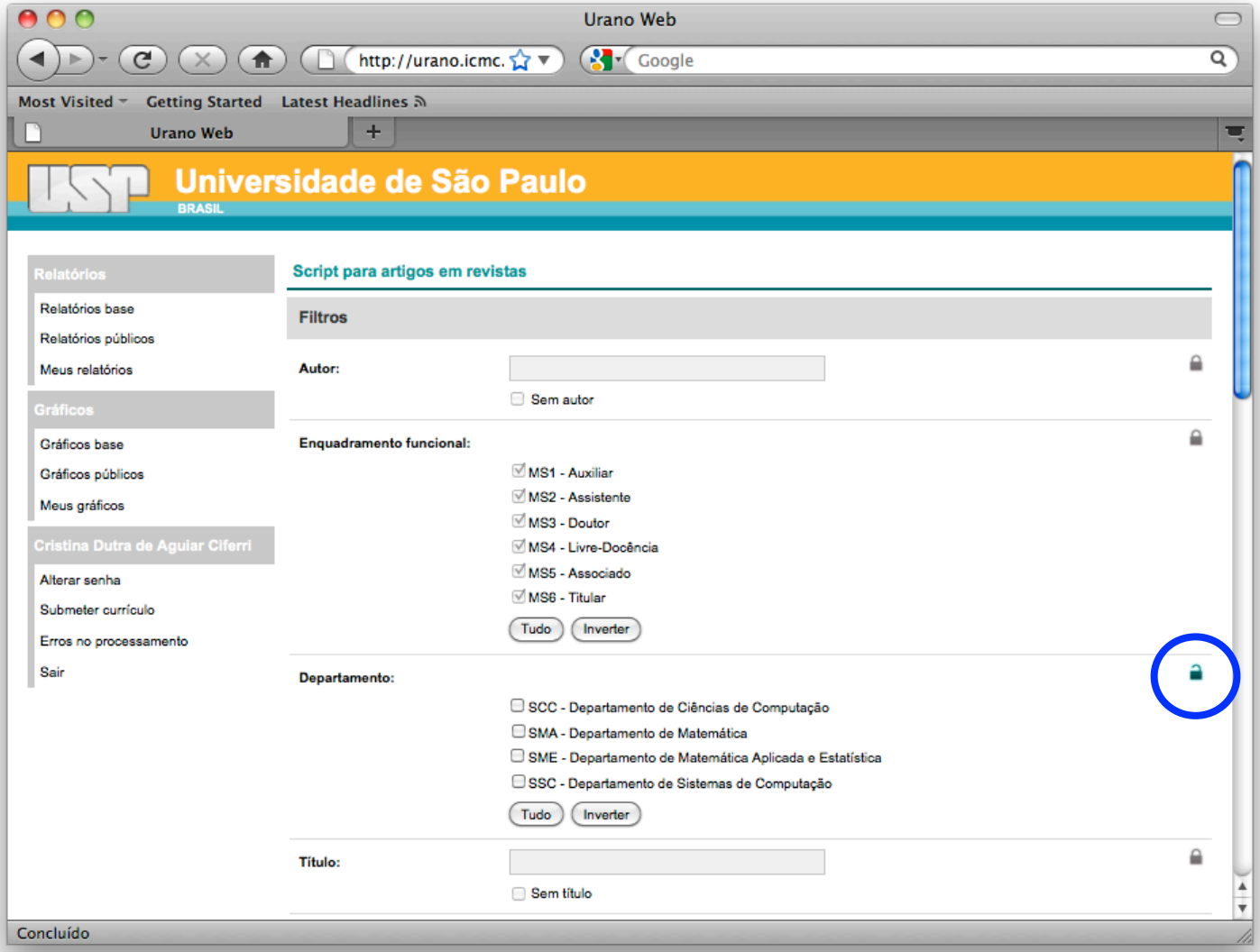
Visão do Investigador



Visão do Investigador



Visão do Pesquisador



Gerenciamento do Log