

Aplicação de Estratégias de Recomendação ao Domínio de Componentes de Software

Tássia Camões Araújo

Universidade de São Paulo

EXAME DE QUALIFICAÇÃO DE MESTRADO

Programa: Ciência da Computação

Orientador: Prof. Dr. Arnaldo Mandel

24 de fevereiro de 2011

Roteiro

- 1 Introdução
- 2 Sistemas de Recomendação
- 3 Distribuições GNU/Linux
- 4 Recomendação nas distribuições
- 5 Desenvolvimento do Trabalho
- 6 Conclusão

Introdução

- Pluralidade de opções
- Infinitude de interesses
- Consumidores produtores: Wikipedia, Flickr, youtube, FOSS
- Sobrecarga gera insatisfação (“mais é menos”)
- Recomendações auxiliam a escolha
- Sistemas recomendadores

Introdução

- Pluralidade de opções
- Infinitude de interesses
- Consumidores produtores: Wikipedia, Flickr, youtube, FOSS
- Sobrecarga gera insatisfação (“mais é menos”)
- Recomendações auxiliam a escolha
- Sistemas recomendadores

Introdução

- Pluralidade de opções
- Infinitude de interesses
- Consumidores produtores: Wikipedia, Flickr, youtube, FOSS
- Sobrecarga gera insatisfação (“mais é menos”)
- Recomendações auxiliam a escolha
- Sistemas recomendadores

Introdução

- Pluralidade de opções
- Infinitude de interesses
- Consumidores produtores: Wikipedia, Flickr, youtube, FOSS
- Sobrecarga gera insatisfação (“mais é menos”)
- Recomendações auxiliam a escolha
- Sistemas recomendadores

Introdução

- Pluralidade de opções
- Infinitude de interesses
- Consumidores produtores: Wikipedia, Flickr, youtube, FOSS
- Sobrecarga gera insatisfação (“mais é menos”)
- Recomendações auxiliam a escolha
- Sistemas recomendadores

Introdução

- Pluralidade de opções
- Infinitude de interesses
- Consumidores produtores: Wikipedia, Flickr, youtube, FOSS
- Sobrecarga gera insatisfação (“mais é menos”)
- Recomendações auxiliam a escolha
- Sistemas recomendadores

Introdução

- Pluralidade de opções
- Infinitude de interesses
- Consumidores produtores: Wikipedia, Flickr, youtube, FOSS
- Sobrecarga gera insatisfação (“mais é menos”)
- Recomendações auxiliam a escolha
- Sistemas recomendadores
- Recomendação de componentes de software

Sistemas de Recomendação

- Ações
- Desafios
- Técnicas
- Estratégias
- Avaliação

Sistemas de Recomendação

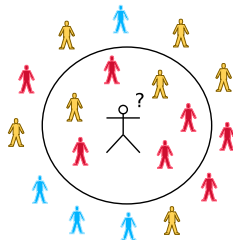
- Ações
- Desafios
- Técnicas
- Estratégias
- Avaliação

Sistemas de Recomendação

- Ações
- Desafios
- Técnicas
- Estratégias
- Avaliação

Técnicas [+]

- K-NN
- Bayes ingênuo
- Tf-idf
- BM25
- Apriori



Técnicas [\[+\]](#)

- K-NN
- Bayes ingênuo
- Tf-idf
- BM25
- Apriori

$$c_{MAP} = \arg \max_{c \in C} \hat{P}(c) \prod_{1 \leq i \leq n} \hat{P}(x_i | c)$$

Técnicas [+]

- K-NN
- Bayes ingênuo
- Tf-idf
- BM25
- Apriori

$$tf-idf_{t,d} = tf_{t,d} \cdot \log \frac{N}{df_t}$$

Técnicas [+]

- K-NN
- Bayes ingênuo
- Tf-idf
- BM25
- Apriori

$$R_{d,q} = \sum_{t \in q} \log \frac{(r + 0.5)(N - n - R + r + 0.5)}{(R - r + 0.5)(n - r + 0.5)} \cdot \frac{tf_{t,D}(k_1 + 1)}{k_1((1 - b) + b \frac{l_d}{l_{avg}}) + tf_{t,D}} \cdot \frac{(k_3 + 1)qtf_{t,Q}}{k_3 + qtf_{t,Q}}$$

-

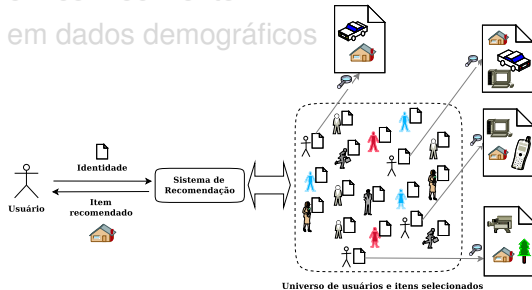
Estratégias de Recomendação

- Reputação dos itens
- Baseada em conteúdo
- Colaborativa
- Baseada em conhecimento
- Baseada em dados demográficos
- Híbrida



Estratégias de Recomendação

- Reputação dos itens
- Baseada em conteúdo
- Colaborativa
- Baseada em conhecimento
- Baseada em dados demográficos
- Híbrida



Estratégias de Recomendação

- Reputação dos itens
- Baseada em conteúdo
- Colaborativa
- Baseada em conhecimento
- Baseada em dados demográficos
- Híbrida

Customers Who Bought This Item Also Bought

 <p>The Structure of Scientific Revolutions by Thomas S. Kuhn</p> <p>★★★★☆ (146)</p> <p>\$7.68</p>	 <p>Conjectures and Refutations: The Growth of Scientific Knowledge by Karl Popper</p> <p>★★★★☆ (11)</p> <p>\$16.42</p>	 <p>Against Method (Fourth Edition) by Paul Feyerabend</p> <p>★★★★☆ (21)</p> <p>\$14.49</p>	 <p>The Open Society and Its Enemies, Vol. I by Sir Karl Raimund Popper</p> <p>★★★★☆ (88)</p> <p>\$19.77</p>
--	---	---	--

Estratégias de Recomendação

- Reputação dos itens
- Baseada em conteúdo
- Colaborativa
- Baseada em conhecimento
- Baseada em dados demográficos
- Híbrida

Estratégias de Recomendação

- Reputação dos itens
- Baseada em conteúdo
- Colaborativa
- Baseada em conhecimento
- Baseada em dados demográficos
- Híbrida
 - ponderação, revezamento, combinação, cascata, combinação ou acréscimo de atributos e meta-nível

Estratégias de Recomendação

- Reputação dos itens
- Baseada em conteúdo
- Colaborativa
- Baseada em conhecimento
- Baseada em dados demográficos
- Híbrida
 - ponderação, revezamento, combinação, cascata, combinação ou acréscimo de atributos e meta-nível

Avaliação de Recomendadores [+]

- Ações como guia de metodologia
- Seleção dos dados
 - *Análise offline vs experimentos "ao vivo"*
 - *Dados reais vs sintéticos*
- Métricas
 - *Acurácia de classificação*
 - *Acurácia de predição*
 - *Cobertura, curva de aprendizado, novidade, surpresa*
 - *Performance*

Avaliação de Recomendadores [+]

- Ações como guia de metodologia
- Seleção dos dados
 - Análise *offline* vs experimentos “*ao vivo*”
 - Dados reais vs sintéticos
- Métricas
 - Acurácia de classificação
 - Acurácia de predição
 - Cobertura, curva de aprendizado, novidade, surpresa
 - Performance

Avaliação de Recomendadores [+]

- Ações como guia de metodologia
- Seleção dos dados
 - Análise *offline* vs experimentos “ao vivo”
 - Dados reais vs sintéticos
- Métricas
 - Acurácia de classificação
 - Acurácia de predição
 - Cobertura, curva de aprendizado, novidade, surpresa
 - Performance

Avaliação de Recomendadores [+]

- Ações como guia de metodologia
- Seleção dos dados
 - Análise *offline* vs experimentos “*ao vivo*”
 - Dados reais vs sintéticos
- Métricas
 - Acurácia de classificação
 - Acurácia de predição
 - Cobertura, curva de aprendizado, novidade, surpresa
 - Performance

Avaliação de Recomendadores [+]

- Ações como guia de metodologia
- Seleção dos dados
 - Análise *offline* vs experimentos “ao vivo”
 - Dados reais vs sintéticos
- Métricas
 - Acurácia de classificação
 - Acurácia de predição
 - Cobertura, curva de aprendizado, novidade, surpresa
 - Performance

Avaliação de Recomendadores [+]

- Ações como guia de metodologia
- Seleção dos dados
 - Análise *offline* vs experimentos “*ao vivo*”
 - Dados reais vs sintéticos
- Métricas
 - Acurácia de classificação
 - Acurácia de predição
 - Cobertura, curva de aprendizado, novidade, surpresa
 - Performance

Avaliação de Recomendadores [+]

- Ações como guia de metodologia
- Seleção dos dados
 - Análise *offline* vs experimentos “*ao vivo*”
 - Dados reais vs sintéticos
- Métricas
 - Acurácia de classificação
 - Acurácia de predição
 - Cobertura, curva de aprendizado, novidade, surpresa
 - Performance

Avaliação de Recomendadores [+]

- Ações como guia de metodologia
- Seleção dos dados
 - Análise *offline* vs experimentos “*ao vivo*”
 - Dados reais vs sintéticos
- Métricas
 - Acurácia de classificação
 - Acurácia de predição
 - Cobertura, curva de aprendizado, novidade, surpresa
 - Performance

Avaliação de Recomendadores [+]

- Ações como guia de metodologia
- Seleção dos dados
 - Análise *offline* vs experimentos “ao vivo”
 - Dados reais vs sintéticos
- Métricas
 - Acurácia de classificação
 - Acurácia de predição
 - Cobertura, curva de aprendizado, novidade, surpresa
 - Performance

Distribuições GNU/Linux

- Aplicativos GNU + kernel Linux
- Encapsulamento de componentes em pacotes
- Seleção e configuração de aplicativos
- Escolha do Debian GNU/Linux como plataforma
 - Esquema consistente de distribuição de aplicativos: APT
 - Disponibilidade de dados estatísticos: Popcon
 - Possibilidade de integração dos resultados
 - Popularidade

Distribuições GNU/Linux

- Aplicativos GNU + kernel Linux
- Encapsulamento de componentes em pacotes
- Seleção e configuração de aplicativos
- Escolha do Debian GNU/Linux como plataforma
 - Esquema consistente de distribuição de aplicativos: APT
 - Disponibilidade de dados estatísticos: Popcon
 - Possibilidade de integração dos resultados
 - Popularidade

Distribuições GNU/Linux

- Aplicativos GNU + kernel Linux
- Encapsulamento de componentes em pacotes
- Seleção e configuração de aplicativos
- Escolha do Debian GNU/Linux como plataforma
 - Esquema consistente de distribuição de aplicativos: APT
 - Disponibilidade de dados estatísticos: Popcon
 - Possibilidade de integração dos resultados
 - Popularidade

Distribuições GNU/Linux

- Aplicativos GNU + kernel Linux
- Encapsulamento de componentes em pacotes
- Seleção e configuração de aplicativos
- Escolha do Debian GNU/Linux como plataforma
 - Esquema consistente de distribuição de aplicativos: APT
 - Disponibilidade de dados estatísticos: Popcon
 - Possibilidade de integração dos resultados
 - Popularidade

Distribuições GNU/Linux

- Aplicativos GNU + kernel Linux
- Encapsulamento de componentes em pacotes
- Seleção e configuração de aplicativos
- Escolha do Debian GNU/Linux como plataforma
 - Esquema consistente de distribuição de aplicativos: APT
 - Disponibilidade de dados estatísticos: Popcon
 - Possibilidade de integração dos resultados
 - Popularidade

Distribuições GNU/Linux

- Aplicativos GNU + kernel Linux
- Encapsulamento de componentes em pacotes
- Seleção e configuração de aplicativos
- Escolha do Debian GNU/Linux como plataforma
 - Esquema consistente de distribuição de aplicativos: APT
 - Disponibilidade de dados estatísticos: Popcon
 - Possibilidade de integração dos resultados
 - Popularidade

Distribuições GNU/Linux

- Aplicativos GNU + kernel Linux
- Encapsulamento de componentes em pacotes
- Seleção e configuração de aplicativos
- Escolha do Debian GNU/Linux como plataforma
 - Esquema consistente de distribuição de aplicativos: APT
 - Disponibilidade de dados estatísticos: Popcon
 - Possibilidade de integração dos resultados
 - Popularidade

Distribuições GNU/Linux

- Aplicativos GNU + kernel Linux
- Encapsulamento de componentes em pacotes
- Seleção e configuração de aplicativos
- Escolha do Debian GNU/Linux como plataforma
 - Esquema consistente de distribuição de aplicativos: APT
 - Disponibilidade de dados estatísticos: Popcon
 - Possibilidade de integração dos resultados
 - Popularidade

Recomendação nas Distribuições [\[+\]](#)

- **AppInstaller**

- *Cross-distro Meeting on Application Installer (2011)*
- Padrões de instalação e gerenciamento de meta-dados
 - *Open Collaboration Services*
- Destaque para *Debtags* e *Apt-xapian-index*

- Fontes de informação no Debian

- BTS, Popcon, PTS, UDD, Debtags, axi

- Iniciativas descontinuadas

- *PopSuggest* e *Debommender*

Recomendação nas Distribuições [+]

- AppInstaller

- *Cross-distro Meeting on Application Installer (2011)*
- Padrões de instalação e gerenciamento de meta-dados
 - *Open Collaboration Services*
- Destaque para *Debtags* e *Apt-xapian-index*

- Fontes de informação no Debian

- BTS, Popcon, PTS, UDD, Debtags, axi

- Iniciativas descontinuadas

- *PopSuggest* e *Debommender*

Recomendação nas Distribuições [\[+\]](#)

- AppInstaller
 - *Cross-distro Meeting on Application Installer (2011)*
 - Padrões de instalação e gerenciamento de meta-dados
 - *Open Collaboration Services*
 - Destaque para *Debtags* e *Apt-xapian-index*
- Fontes de informação no Debian
 - BTS, Popcon, PTS, UDD, Debtags, axi
- Iniciativas descontinuadas
 - *PopSuggest* e *Debommender*

Recomendação nas Distribuições [+]

- AppInstaller
 - *Cross-distro Meeting on Application Installer (2011)*
 - Padrões de instalação e gerenciamento de meta-dados
 - *Open Collaboration Services*
 - Destaque para *Debtags* e *Apt-xapian-index*
- Fontes de informação no Debian
 - BTS, Popcon, PTS, UDD, Debtags, axi
- Iniciativas descontinuadas
 - *PopSuggest* e *Debommender*

Recomendação nas Distribuições [+]

- AppInstaller
 - *Cross-distro Meeting on Application Installer (2011)*
 - Padrões de instalação e gerenciamento de meta-dados
 - *Open Collaboration Services*
 - Destaque para *Debtags* e *Apt-xapian-index*
- Fontes de informação no Debian
 - BTS, Popcon, PTS, UDD, Debtags, axi
- Iniciativas descontinuadas
 - *PopSuggest* e *Debommender*

Recomendação nas Distribuições [+]

- AppInstaller
 - *Cross-distro Meeting on Application Installer (2011)*
 - Padrões de instalação e gerenciamento de meta-dados
 - *Open Collaboration Services*
 - Destaque para *Debtags* e *Apt-xapian-index*
- Fontes de informação no Debian
 - BTS, Popcon, PTS, UDD, Debtags, axi
- Iniciativas descontinuadas
 - *PopSuggest* e *Debommender*

Recomendação nas Distribuições [+]

- AppInstaller
 - *Cross-distro Meeting on Application Installer (2011)*
 - Padrões de instalação e gerenciamento de meta-dados
 - *Open Collaboration Services*
 - Destaque para *Debtags* e *Apt-xapian-index*
- Fontes de informação no Debian
 - BTS, Popcon, PTS, UDD, Debtags, axi
- Iniciativas descontinuadas
 - *PopSuggest* e *Debommender*

Recomendação nas Distribuições [+]

- AppInstaller
 - *Cross-distro Meeting on Application Installer (2011)*
 - Padrões de instalação e gerenciamento de meta-dados
 - *Open Collaboration Services*
 - Destaque para *Debtags* e *Apt-xapian-index*
- Fontes de informação no Debian
 - BTS, Popcon, PTS, UDD, Debtags, axi
- Iniciativas descontinuadas
 - *PopSuggest* e *Debommender*

Recomendação nas Distribuições [+]

- AppInstaller
 - *Cross-distro Meeting on Application Installer (2011)*
 - Padrões de instalação e gerenciamento de meta-dados
 - *Open Collaboration Services*
 - Destaque para *Debtags* e *Apt-xapian-index*
- Fontes de informação no Debian
 - BTS, Popcon, PTS, UDD, Debtags, axi
- Iniciativas descontinuadas
 - *PopSuggest* e *Debommender*

Desenvolvimento do Trabalho

- Caracterização do problema
- Metodologia
- Plano de execução

Caracterização do Problema

- Modelagem de clientes e itens
- Ação do recomendador: encontrar itens relevantes
- Pontuação binária vs multi-valorada
- Sobre dependências:
 - Na mesma recomendação
 - Em associações
- Sobre necessidade dos usuários:
 - Modelagem como funcionalidades
 - Pacotes virtuais e tags

Caracterização do Problema

- Modelagem de clientes e itens
- Ação do recomendador: encontrar itens relevantes
- Pontuação binária vs multi-valorada
- Sobre dependências:
 - Na mesma recomendação
 - Em associações
- Sobre necessidade dos usuários:
 - Modelagem como funcionalidades
 - Pacotes virtuais e tags

Caracterização do Problema

- Modelagem de clientes e itens
- Ação do recomendador: encontrar itens relevantes
- Pontuação binária vs multi-valorada
- Sobre dependências:
 - Na mesma recomendação
 - Em associações
- Sobre necessidade dos usuários:
 - Modelagem como funcionalidades
 - Pacotes virtuais e tags

Caracterização do Problema

- Modelagem de clientes e itens
- Ação do recomendador: encontrar itens relevantes
- Pontuação binária vs multi-valorada
- Sobre dependências:
 - Na mesma recomendação
 - Em associações
- Sobre necessidade dos usuários:
 - Modelagem como funcionalidades
 - Pacotes virtuais e tags

Caracterização do Problema

- Modelagem de clientes e itens
- Ação do recomendador: encontrar itens relevantes
- Pontuação binária vs multi-valorada
- Sobre dependências:
 - Na mesma recomendação
 - Em associações
- Sobre necessidade dos usuários:
 - Modelagem como funcionalidades
 - Pacotes virtuais e tags

Caracterização do Problema

- Modelagem de clientes e itens
- Ação do recomendador: encontrar itens relevantes
- Pontuação binária vs multi-valorada
- Sobre dependências:
 - Na mesma recomendação
 - Em associações
- Sobre necessidade dos usuários:
 - Modelagem como funcionalidades
 - Pacotes virtuais e tags

Caracterização do Problema

- Modelagem de clientes e itens
- Ação do recomendador: encontrar itens relevantes
- Pontuação binária vs multi-valorada
- Sobre dependências:
 - Na mesma recomendação
 - Em associações
- Sobre necessidade dos usuários:
 - Modelagem como funcionalidades
 - Pacotes virtuais e tags

Caracterização do Problema

- Modelagem de clientes e itens
- Ação do recomendador: encontrar itens relevantes
- Pontuação binária vs multi-valorada
- Sobre dependências:
 - Na mesma recomendação
 - Em associações
- Sobre necessidade dos usuários:
 - Modelagem como funcionalidades
 - Pacotes virtuais e tags

Caracterização do Problema

- Modelagem de clientes e itens
- Ação do recomendador: encontrar itens relevantes
- Pontuação binária vs multi-valorada
- Sobre dependências:
 - Na mesma recomendação
 - Em associações
- Sobre necessidade dos usuários:
 - Modelagem como funcionalidades
 - Pacotes virtuais e tags

Metodologia

- Coleta de dados
 - *Popcon, UDD e Debtags*
- Seleção de atributos
 - Instalação padrão
 - Dependência entre pacotes
 - Dados temporais
- Estratégias de recomendação
- Codificação
- Avaliação

Metodologia

- Coleta de dados
 - *Popcon, UDD e Debtags*
- Seleção de atributos
 - Instalação padrão
 - Dependência entre pacotes
 - Dados temporais
- Estratégias de recomendação
- Codificação
- Avaliação

Metodologia

- Coleta de dados
 - *Popcon, UDD e Debtags*
- Seleção de atributos
 - Instalação padrão
 - Dependência entre pacotes
 - Dados temporais
- Estratégias de recomendação
- Codificação
- Avaliação

Metodologia

- Coleta de dados
 - *Popcon, UDD e Debtags*
- Seleção de atributos
 - Instalação padrão
 - Dependência entre pacotes
 - Dados temporais
- Estratégias de recomendação
- Codificação
- Avaliação

Metodologia

- Coleta de dados
 - *Popcon, UDD e Debtags*
- Seleção de atributos
 - Instalação padrão
 - Dependência entre pacotes
 - Dados temporais
- Estratégias de recomendação
- Codificação
- Avaliação

Metodologia

- Coleta de dados
 - *Popcon, UDD e Debtags*
- Seleção de atributos
 - Instalação padrão
 - Dependência entre pacotes
 - Dados temporais
- Estratégias de recomendação
- Codificação
- Avaliação

Metodologia

Estratégias de recomendação

- Baseada em conteúdo

- Técnicas: *BM25* e *Bayes* ingênuo
- Dados: *tags* e descrições dos pacotes

- Colaborativa

- Técnica: *K-NN* e *K-means*
- Dados: listas de pacotes instalados

- Híbrida

- Cascata: refinamento de resultados a partir de dados demográficos e reputação dos itens
- Meta-nível: colaboração por conteúdo
- Revezamento: decisão a partir da seleção de atributos
- Combinação: sugestão de múltiplos recomendadores

Metodologia

Estratégias de recomendação

- Baseada em conteúdo
 - Técnicas: *BM25* e *Bayes ingênuo*
 - Dados: *tags* e descrições dos pacotes
- Colaborativa
 - Técnica: *K-NN* e *K-means*
 - Dados: listas de pacotes instalados
- Híbrida
 - Cascata: refinamento de resultados a partir de dados demográficos e reputação dos itens
 - Meta-nível: colaboração por conteúdo
 - Revezamento: decisão a partir da seleção de atributos
 - Combinação: sugestão de múltiplos recomendadores

Metodologia

Estratégias de recomendação

- Baseada em conteúdo
 - Técnicas: *BM25* e *Bayes* ingênuo
 - Dados: *tags* e descrições dos pacotes
- Colaborativa
 - Técnica: *K-NN* e *K-means*
 - Dados: listas de pacotes instalados
- Híbrida
 - Cascata: refinamento de resultados a partir de dados demográficos e reputação dos itens
 - Meta-nível: colaboração por conteúdo
 - Revezamento: decisão a partir da seleção de atributos
 - Combinação: sugestão de múltiplos recomendadores

Metodologia

Estratégias de recomendação

- Baseada em conteúdo
 - Técnicas: *BM25* e *Bayes* ingênuo
 - Dados: *tags* e descrições dos pacotes
- Colaborativa
 - Técnica: *K-NN* e *K-means*
 - Dados: listas de pacotes instalados
- Híbrida
 - Cascata: refinamento de resultados a partir de dados demográficos e reputação dos itens
 - Meta-nível: colaboração por conteúdo
 - Revezamento: decisão a partir da seleção de atributos
 - Combinação: sugestão de múltiplos recomendadores

Metodologia

Estratégias de recomendação

- Baseada em conteúdo
 - Técnicas: *BM25* e *Bayes* ingênuo
 - Dados: *tags* e descrições dos pacotes
- Colaborativa
 - Técnica: *K-NN* e *K-means*
 - Dados: listas de pacotes instalados
- Híbrida
 - Cascata: refinamento de resultados a partir de dados demográficos e reputação dos itens
 - Meta-nível: colaboração por conteúdo
 - Revezamento: decisão a partir da seleção de atributos
 - Combinação: sugestão de múltiplos recomendadores

Metodologia

Estratégias de recomendação

- Baseada em conteúdo
 - Técnicas: *BM25* e *Bayes* ingênuo
 - Dados: *tags* e descrições dos pacotes
- Colaborativa
 - Técnica: *K-NN* e *K-means*
 - Dados: listas de pacotes instalados
- Híbrida
 - Cascata: refinamento de resultados a partir de dados demográficos e reputação dos itens
 - Meta-nível: colaboração por conteúdo
 - Revezamento: decisão a partir da seleção de atributos
 - Combinação: sugestão de múltiplos recomendadores

Metodologia

Estratégias de recomendação

- Baseada em conteúdo
 - Técnicas: *BM25* e *Bayes* ingênuo
 - Dados: *tags* e descrições dos pacotes
- Colaborativa
 - Técnica: *K-NN* e *K-means*
 - Dados: listas de pacotes instalados
- Híbrida
 - Cascata: refinamento de resultados a partir de dados demográficos e reputação dos itens
 - Meta-nível: colaboração por conteúdo
 - Revezamento: decisão a partir da seleção de atributos
 - Combinação: sugestão de múltiplos recomendadores

Metodologia

Estratégias de recomendação

- Baseada em conteúdo
 - Técnicas: *BM25* e *Bayes* ingênuo
 - Dados: *tags* e descrições dos pacotes
- Colaborativa
 - Técnica: *K-NN* e *K-means*
 - Dados: listas de pacotes instalados
- Híbrida
 - Cascata: refinamento de resultados a partir de dados demográficos e reputação dos itens
 - Meta-nível: colaboração por conteúdo
 - Revezamento: decisão a partir da seleção de atributos
 - Combinação: sugestão de múltiplos recomendadores

Metodologia

Estratégias de recomendação

- Baseada em conteúdo
 - Técnicas: *BM25* e *Bayes* ingênuo
 - Dados: *tags* e descrições dos pacotes
- Colaborativa
 - Técnica: *K-NN* e *K-means*
 - Dados: listas de pacotes instalados
- Híbrida
 - Cascata: refinamento de resultados a partir de dados demográficos e reputação dos itens
 - Meta-nível: colaboração por conteúdo
 - Revezamento: decisão a partir da seleção de atributos
 - Combinação: sugestão de múltiplos recomendadores

Metodologia

Estratégias de recomendação

- Baseada em conteúdo
 - Técnicas: *BM25* e *Bayes* ingênuo
 - Dados: *tags* e descrições dos pacotes
- Colaborativa
 - Técnica: *K-NN* e *K-means*
 - Dados: listas de pacotes instalados
- Híbrida
 - Cascata: refinamento de resultados a partir de dados demográficos e reputação dos itens
 - Meta-nível: colaboração por conteúdo
 - Revezamento: decisão a partir da seleção de atributos
 - Combinação: sugestão de múltiplos recomendadores

Metodologia

Estratégias de recomendação

- Baseada em conteúdo
 - Técnicas: *BM25* e *Bayes* ingênuo
 - Dados: *tags* e descrições dos pacotes
- Colaborativa
 - Técnica: *K-NN* e *K-means*
 - Dados: listas de pacotes instalados
- Híbrida
 - Cascata: refinamento de resultados a partir de dados demográficos e reputação dos itens
 - Meta-nível: colaboração por conteúdo
 - Revezamento: decisão a partir da seleção de atributos
 - Combinação: sugestão de múltiplos recomendadores

Metodologia

Codificação

- Linguagem *Python*
- Bibliotecas: *Xapian* e *NLTK*
- Documentação: *Doxygen*
- Controle de versão: *Git*

Metodologia

Codificação

- Linguagem *Python*
- Bibliotecas: *Xapian* e *NLTK*
- Documentação: *Doxygen*
- Controle de versão: *Git*

Metodologia

Codificação

- Linguagem *Python*
- Bibliotecas: *Xapian* e *NLTK*
- Documentação: *Doxygen*
- Controle de versão: *Git*

Metodologia

Codificação

- Linguagem *Python*
- Bibliotecas: *Xapian* e *NLTK*
- Documentação: *Doxygen*
- Controle de versão: *Git*

Metodologia

Avaliação

- Testes preliminares: validação cruzada
- Seleção de estratégias e ajustes das soluções
- Consulta pública: *survey* eletrônico
 - Composição da identidade do usuário
 - Produção das recomendações por diversas estratégias
 - Apresentação dos resultados
 - Avaliação pelo usuário
 - Aplicação de métricas
- Compilação dos resultados globais

Metodologia

Avaliação

- Testes preliminares: validação cruzada
- Seleção de estratégias e ajustes das soluções
- Consulta pública: *survey* eletrônico
 - Composição da identidade do usuário
 - Produção das recomendações por diversas estratégias
 - Apresentação dos resultados
 - Avaliação pelo usuário
 - Aplicação de métricas
- Compilação dos resultados globais

Metodologia

Avaliação

- Testes preliminares: validação cruzada
- Seleção de estratégias e ajustes das soluções
- Consulta pública: *survey* eletrônico
 - 1 Composição da identidade do usuário
 - 2 Produção das recomendações por diversas estratégias
 - 3 Apresentação dos resultados
 - 4 Avaliação pelo usuário
 - 5 Aplicação de métricas
- Compilação dos resultados globais

Metodologia

Avaliação

- Testes preliminares: validação cruzada
- Seleção de estratégias e ajustes das soluções
- Consulta pública: *survey* eletrônico
 - 1 Composição da identidade do usuário
 - 2 Produção das recomendações por diversas estratégias
 - 3 Apresentação dos resultados
 - 4 Avaliação pelo usuário
 - 5 Aplicação de métricas
- Compilação dos resultados globais

Metodologia

Avaliação

- Testes preliminares: validação cruzada
- Seleção de estratégias e ajustes das soluções
- Consulta pública: *survey* eletrônico
 - 1 Composição da identidade do usuário
 - 2 Produção das recomendações por diversas estratégias
 - 3 Apresentação dos resultados
 - 4 Avaliação pelo usuário
 - 5 Aplicação de métricas
- Compilação dos resultados globais

Metodologia

Avaliação

- Testes preliminares: validação cruzada
- Seleção de estratégias e ajustes das soluções
- Consulta pública: *survey* eletrônico
 - 1 Composição da identidade do usuário
 - 2 Produção das recomendações por diversas estratégias
 - 3 Apresentação dos resultados
 - 4 Avaliação pelo usuário
 - 5 Aplicação de métricas
- Compilação dos resultados globais

Metodologia

Avaliação

- Testes preliminares: validação cruzada
- Seleção de estratégias e ajustes das soluções
- Consulta pública: *survey* eletrônico
 - 1 Composição da identidade do usuário
 - 2 Produção das recomendações por diversas estratégias
 - 3 Apresentação dos resultados
 - 4 Avaliação pelo usuário
 - 5 Aplicação de métricas
- Compilação dos resultados globais

Metodologia

Avaliação

- Testes preliminares: validação cruzada
- Seleção de estratégias e ajustes das soluções
- Consulta pública: *survey* eletrônico
 - 1 Composição da identidade do usuário
 - 2 Produção das recomendações por diversas estratégias
 - 3 Apresentação dos resultados
 - 4 Avaliação pelo usuário
 - 5 Aplicação de métricas
- Compilação dos resultados globais

Metodologia

Avaliação

- Testes preliminares: validação cruzada
- Seleção de estratégias e ajustes das soluções
- Consulta pública: *survey* eletrônico
 - 1 Composição da identidade do usuário
 - 2 Produção das recomendações por diversas estratégias
 - 3 Apresentação dos resultados
 - 4 Avaliação pelo usuário
 - 5 Aplicação de métricas
- Compilação dos resultados globais

Plano de Execução

Atividade	Jan	Fev	Mar	Abr	Mai	Jun	Jul
Preparação e realização da qualificação	✓	✓					
Implementação de diferentes estratégias e técnicas		✓	X	X			
Testes e ajustes na implementação				X	X		
Preparação e aplicação do <i>survey</i>					X	X	
Escrita da dissertação					X	X	X

Eventos: FISL12 (jun), Debconf11 (jul) e OSS2011 (out)

Conclusão

- Acompanhamento das discussões no *FreeDesktop*
- Conformidade com padrões
- Conclusão pautada pelos resultados do *survey*
- Meta++: integração de um recomendador na estrutura do *software-center*

Conclusão

- Acompanhamento das discussões no *FreeDesktop*
- Conformidade com padrões
- Conclusão pautada pelos resultados do *survey*
- Meta++: integração de um recomendador na estrutura do *software-center*

Conclusão

- Acompanhamento das discussões no *FreeDesktop*
- Conformidade com padrões
- Conclusão pautada pelos resultados do *survey*
- Meta++: integração de um recomendador na estrutura do *software-center*

Conclusão

- Acompanhamento das discussões no *FreeDesktop*
- Conformidade com padrões
- Conclusão pautada pelos resultados do *survey*
- Meta++: integração de um recomendador na estrutura do *software-center*

Referências

- Araújo, T. C. *Aplicação de Estratégias de Recomendação ao Domínio de Componentes de Software*. 2011. Disponível em <https://github.com/tassia/QualificacaoMestrado>
- Araújo, T. C. *AppRecommender: Application Recommender for GNU/Linux Systems*. 2011. Disponível em <https://github.com/tassia/AppRecommender>