



UNIVERSIDADE FEDERAL DE PERNAMBUCO  
CENTRO DE INFORMÁTICA  
IN1152 – RECUPERAÇÃO INTELIGENTE DE INFORMAÇÃO



# SISTEMAS DE RECOMENDAÇÃO

Alysson Bispo, Larissa Falcão, Micael Soares, Samuel Romeiro  
{abp2, lctf, msf3, scras}@cin.ufpe.br

# AGENDA

- Introdução
- Definições
- Principais Funções;
- Principais Técnicas de Sistemas de Recomendação:
  - Data Mining;
  - Content-Based;
  - Collaborative Filtering;
  - Context-Aware;
- Avaliação de SR's;
- Sistemas de Recomendação em Redes Sociais;



# INTRODUÇÃO

- No dia-a-dia confiamos em sugestões ao realizar escolhas;
  - Ao comprar produtos, contratar serviços e até se relacionar com pessoas;
  - Fontes de recomendação: Amigos, familiares, profissionais especializados, críticas em jornais/revistas/sites, etc...
- **Motivo?**
  - Grande variedade de opções
  - Dificuldade em filtrar para encontrar o que nos satisfaz

# INTRODUÇÃO

- O que é:  
*Sistemas de Recomendação (SR) são ferramentas ou técnicas capazes de sugerir itens a serem utilizados por um determinado usuário;*
- (***normalmente***) Focados em um tipo de item;
  - Prover informações úteis e efetivas sobre o item em questão;
  - Usuários (ou grupos de usuários) recebem recomendações diferentes - ***personalizado***;

# INTRODUÇÃO

- No início, SR eram realizados por observações e recomendações dos próprios usuários;
- **Abordagem Colaborativa:**  
*“(...) Escolha é bom. Muitas escolhas nem sempre é melhor (...)”*
- Necessidade por recomendações personalizadas:
  - E-commerce;

# INTRODUÇÃO

- Atualmente:
  - SR gera recomendações através de vários tipos de informações e dados sobre o usuário;
  - O próprio usuário alimenta o sistema (feedbacks);
- Área relativamente nova (anos 90..);

# DEFINIÇÕES

- **ITENS:**  
*“(...) São os objetos que são recomendados. Os itens podem ser caracterizadas pela sua complexidade e seu valor ou utilidade. O valor de um item pode ser positivo se o item é útil para o usuário, ou negativo se o item não é adequada e que o usuário tomou uma decisão errada ao escolhê-lo (...)”*
- **USUÁRIOS:**  
*“(...) usuários podem ter objetivos e características muito diversas. A fim de personalizar as recomendações e a interação humano-computador, SR exploram uma vasta gama de informações sobre esses usuários (...)”*
- **TRANSAÇÕES:**  
*“(...) Transações são interações entre um usuário e um SR. As transações são dados armazenados em Log, onde informações importantes são geradas durante a interação humano-computador e que são úteis para o algoritmo de geração de recomendação que o sistema está usando (...)”*

# PRINCIPAIS FUNÇÕES

- LOCALIZAR UM CONJUNTO DE ITENS POTENCIALMENTE RELEVANTES A UM DETERMINADO USUÁRIO;
- E suas tarefas?
  - Achar ALGUNS itens;
  - Achar TODOS os itens;
  - Anotação em contexto;
  - Recomendação de sequências;
  - Recomendar pacotes;
  - *Entre outros....*



# IMPORTÂNCIA

- **Aumentar o número de itens vendidos**
  - Ao invés de comprar X, o usuário compra X, Y e Z
- **Aumentar a diversidade de itens vendidos**
- **Aumentar a satisfação do usuário**
- **Aumentar a fidelidade do usuário**
  - Recomendações são refinadas com o uso
- **Entender melhor o que o usuário quer**
  - As preferências dos usuários podem ser utilizada para outros fins, como melhorar o gerenciamento de estoque

# PRINCIPAIS ABORDAGENS DE SR's

- Content-Based;
- Collaborative Filtering;
- Demográficos;
- Knowledge-Based;
- Community-Based;
- Sistemas de Recomendação Híbridos;

# PRINCIPAIS ABORDAGENS DE SR's

- **Content-Based;**
- **Collaborative Filtering;**
- Demográficos;
- Knowledge-Based;
- Community-Based;
- Sistemas de Recomendação Híbridos;

# DATA MINING



# MINERAÇÃO DE DADOS- DEFINIÇÃO

- Sinônimo para o processo de KDD – *Knowledge Discovery in Databases*;
  - Processo de identificação de padrões potencialmente úteis em base de dados;
- Área multidisciplinar (IA, BD, estatística);

# MINERAÇÃO DE DADOS - ETAPAS

## Pré-processamento de dados

- Medidas de similaridade;
- Amostragem;

## Análise

- Classificação
- Regras de Associação
- Agrupamento;

## Interpretação

# MINERAÇÃO DE DADOS - ETAPAS

## Pré-processamento de dados

- Medidas de similaridade;
- Amostragem;

## Análise

- Classificação
- Regras de Associação
- Agrupamento;

## Interpretação

# MINERAÇÃO DE DADOS – PRÉ-PROCESSAMENTO

- Medidas de similaridade – maior parte dos sistemas de classificação e agrupamento dependem de técnicas apropriadas de similaridade;
  - Distância Euclidiana;
  - Distância de Minkowski;
  - Distância de Mahalanobis;
  - Distância do cosseno;



# MINERAÇÃO DE DADOS – PRÉ-PROCESSAMENTO

- Amostragem – seleção de um subconjunto dos dados de um grande conjunto de dados;
- Amostragem pode ser usado porque o processamento de todo o conjunto de dados é muitas vezes computacionalmente muito caro

# MINERAÇÃO DE DADOS - ETAPAS

## Pré-processamento de dados

- Medidas de similaridade;
- Amostragem;

## Análise

- Classificação
- Regras de Associação
- Agrupamento;

## Interpretação

# MINERAÇÃO DE DADOS – ANÁLISE

- Classificação – analisa características de um registro (objeto) para enquadrá-lo (classificá-lo) em um conjunto pré-definido (classe);
  - Utiliza um conjunto de treinamento para obter classes;
- Técnicas:
  - Redes Bayesianas;
  - Redes Neurais;
  - Árvores de decisões;

# MINERAÇÃO DE DADOS – ANÁLISE

- Regras de Associação – Identificação de itens, cuja presença implica na presença de outros itens na mesma “transação”;
  - Informação nova acerca de compra de produtos;
- Aprendizagem não supervisionada + objetivo de descrição
  - Algoritmos: Apriori, FP-Growth, Basic, ECLAT, DIC, entre outros;
- Possuem a notação  $A \rightarrow B$ , onde A e B são conjuntos de itens
- Ex: caso do aumento de vendas da Wal-Mart nas sextas;

# MINERAÇÃO DE DADOS – ANÁLISE

- Agrupamento - procura segmentar populações heterogêneas em subgrupos ou segmentos homogêneos;
  - Os grupos não são pré-definidos (aprendizagem não supervisionada);

Similarity is hard to define, but... “We know it when we see it” Dr. E. Keogh



---

# CONTENT-BASED

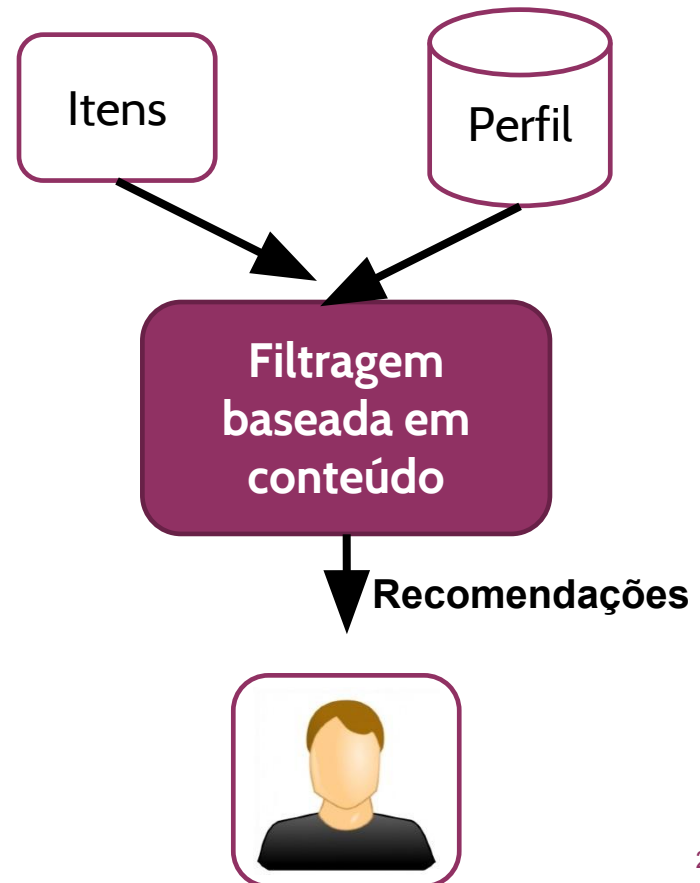


***Técnica que compara a descrição do conteúdo de um determinado item com os interesses dos usuários com a finalidade de verificar se o item é ou não relevante para aquele usuário***

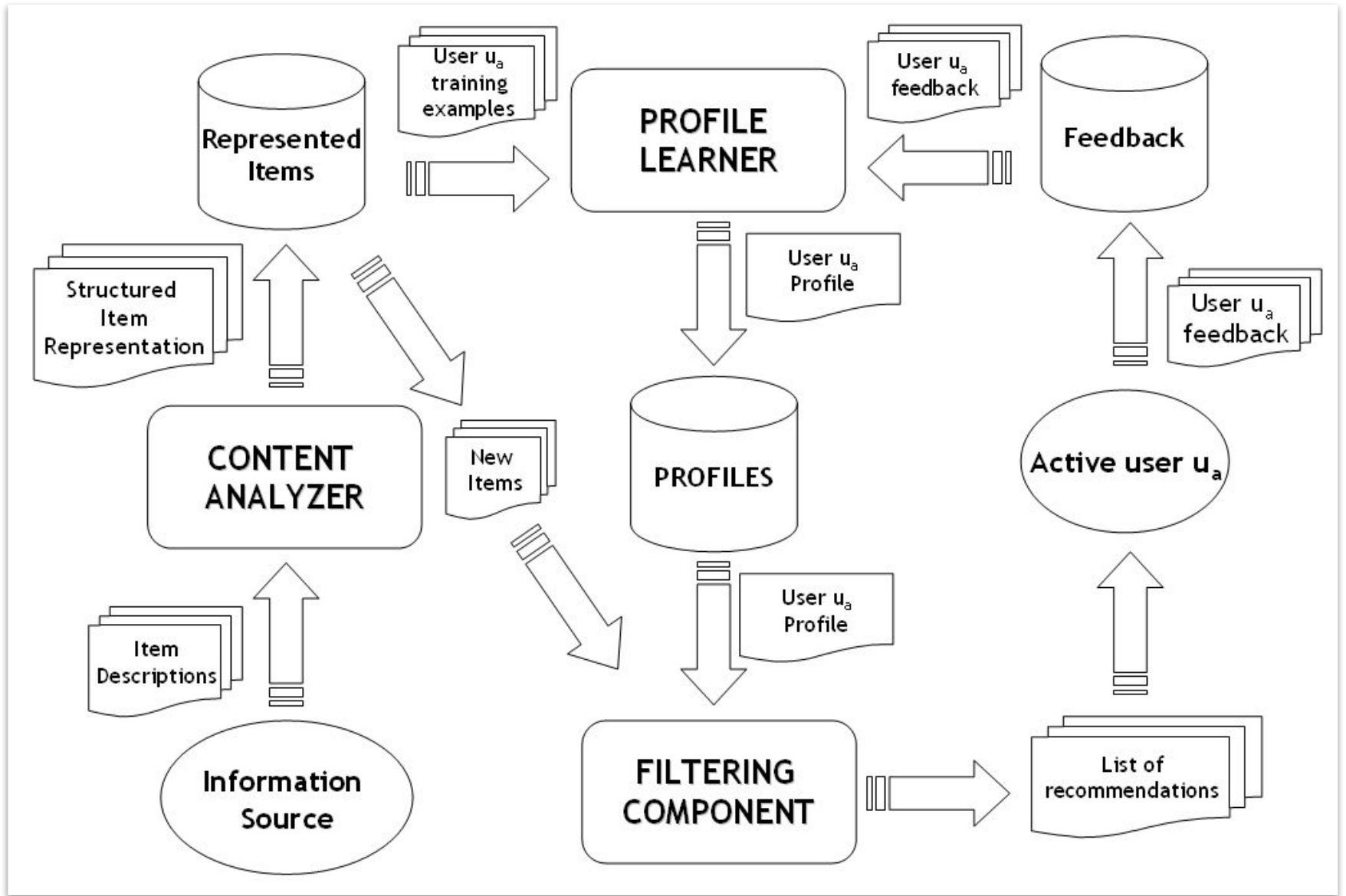


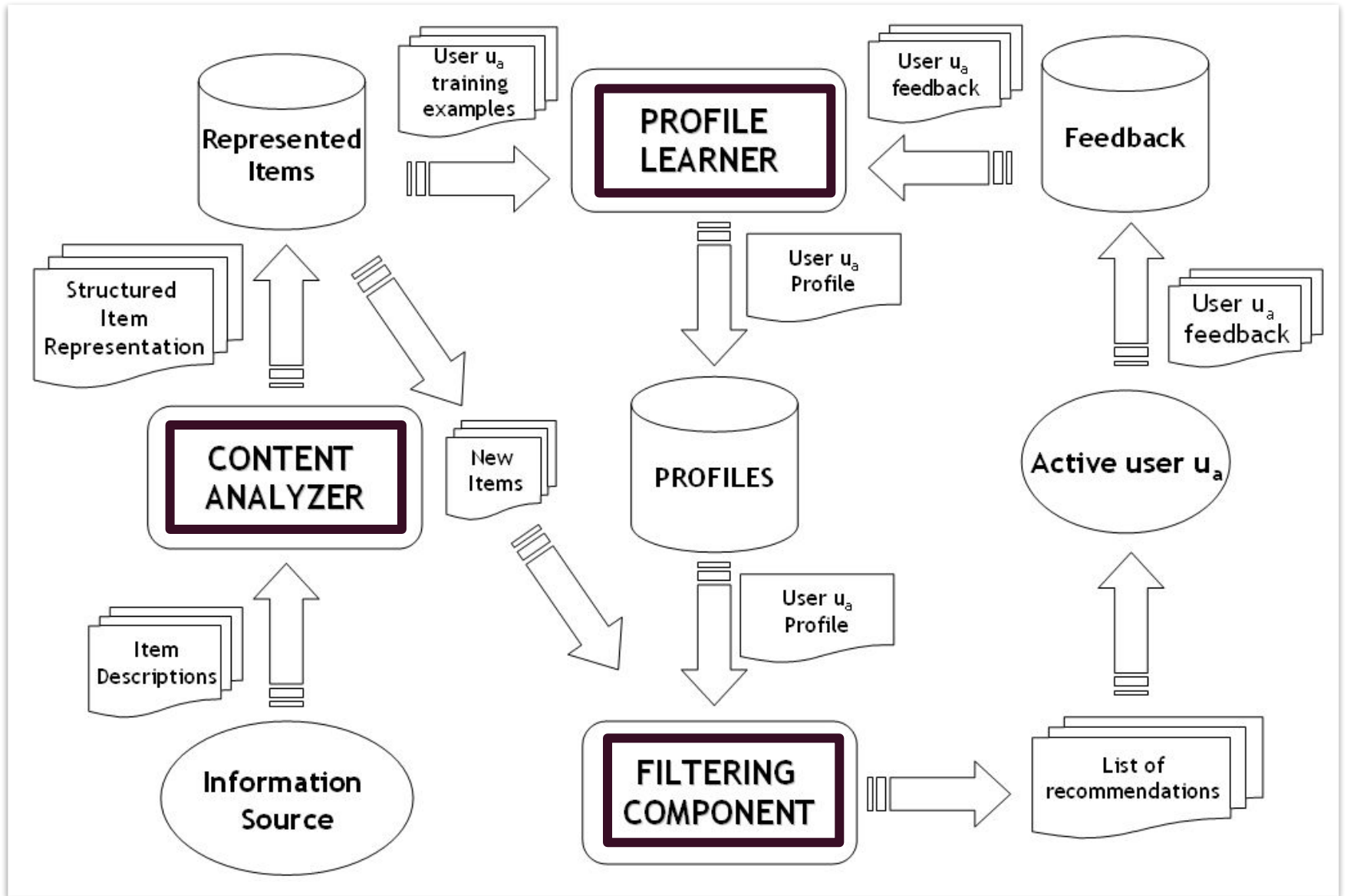
# FILTRAGEM BASEADA EM CONTEÚDO

- O conteúdo normalmente é descrito através de palavras-chave
  - O próprio usuário fornece essa descrição através de suas ações
- Parte do princípio de que os usuários tendem a se interessar por itens similares aos de interesse no passado









# FILTRAGEM BASEADA EM CONTEÚDO: FUNCIONAMENTO GERAL

## Análise de conteúdo

Representar o conteúdo dos itens de forma estruturada

Comum: **Vetor de palavras-chave**

## Construção de perfil

Coletar as preferências do usuário considerando o passado

Normalmente usa técnicas de aprendizagem de máquina

Representação no mesmo formato dos itens

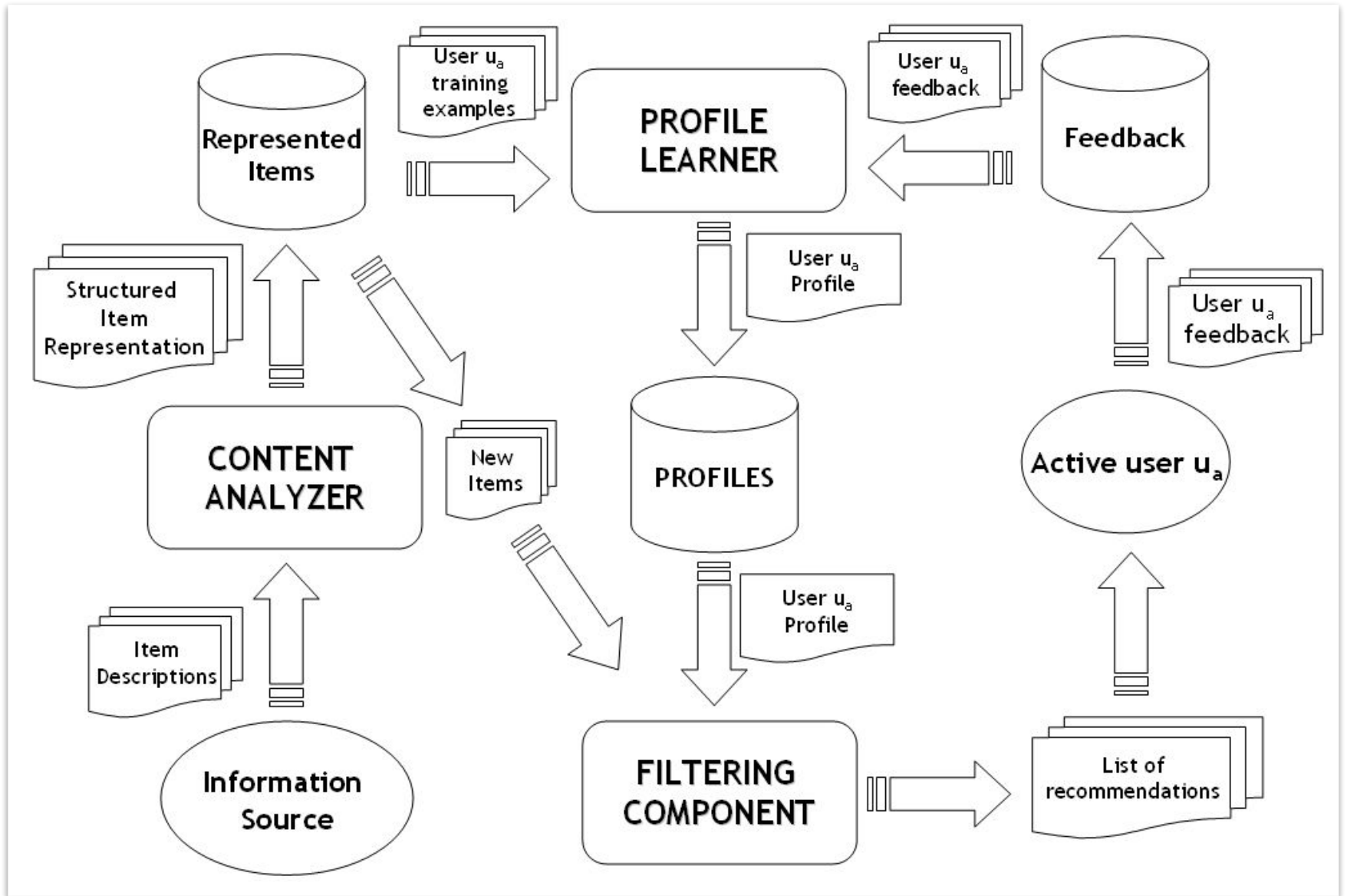
Atualização do perfil a cada nova avaliação

## Filtragem

Fazer a correspondência entre os itens e o perfil do usuário

Resultado: Avaliação de relevância

É usada alguma métrica de **similaridade**



# FILTRAGEM BASEADA EM CONTEÚDO: REPRESENTAÇÃO DE CONTEÚDO

- **Indexação de frequência de termos**
  - Informações dos documentos e interesses dos usuários são descritas por vetores com uma dimensão para cada palavra
  - Cada componente do vetor é a frequência de ocorrência de uma dada palavra na base de dados
  - **TF-IDF**: Atribui um peso maior para palavras que aparecem muito em um documento, mas que aparecem em poucos documentos

# FILTRAGEM BASEADA EM CONTEÚDO: REPRESENTAÇÃO

- Pouco utilizados:
  - Índice de busca booleana: a consulta constitui-se em um conjunto de palavras-chave unidas por operadores booleanos;
  - Filtragem probabilística: raciocínio probabilístico é aplicado para determinar a probabilidade que um documento possui de atender as necessidades de informação de um usuário;

# FILTRAGEM BASEADA EM CONTEÚDO: VANTAGENS

- **Independência do usuário:**
  - Construção de um perfil de usuário baseado unicamente em suas preferências;
- **Transparência:**
  - É possível verificar a razão pela qual um determinado item foi recomendado (listando explicitamente recursos de conteúdo ou descrições que resultaram naquele item);
- **Novo item:**
  - Possibilidade de recomendar itens que ainda não foram avaliados por nenhum usuário;

# FILTRAGEM BASEADA EM CONTEÚDO: LIMITAÇÕES

- **Análise de conteúdo é limitada**
  - É preciso ter uma descrição rica de conteúdo;
  - Lidar com dados pouco estruturados (imagem e áudio) é difícil;
  - No caso de texto, não é possível distinguir um artigo bem escrito de um mal escrito se eles utilizam termos muito semelhantes;
- **Super especialização:**
  - O usuário não receberá indicações do que nunca “consumiu”;
- **Novo usuário:**
  - Um número considerável de avaliações deve ser coletadas antes do sistema começar a fornecer recomendações confiáveis a novos usuários;



# COLLABORATIVE FILTERING



# Introdução

- Sistemas de recomendação dependem de vários tipos de entrada.
- O mais conveniente é o feedback explícito de alta qualidade, onde os usuários reportam diretamente sobre o seu interesse em produtos.
- Mas o feedback explícito nem sempre está disponível, então alguns recommenders inferem as preferências do usuário a partir do feedback implícito mais abundante.

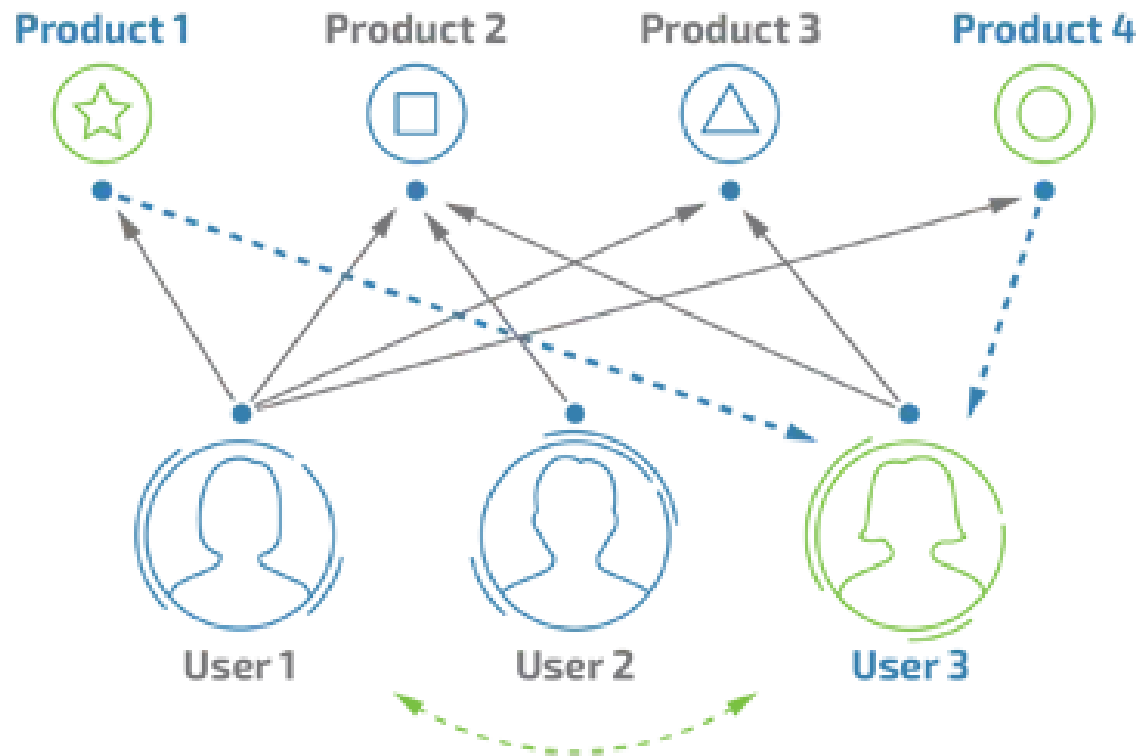
## Definição

**São métodos que produzem recomendações específicas para usuários de itens com base em padrões de classificações ou de uso (por exemplo, compras), sem necessidade de informação exógena sobre qualquer itens ou usuários.**

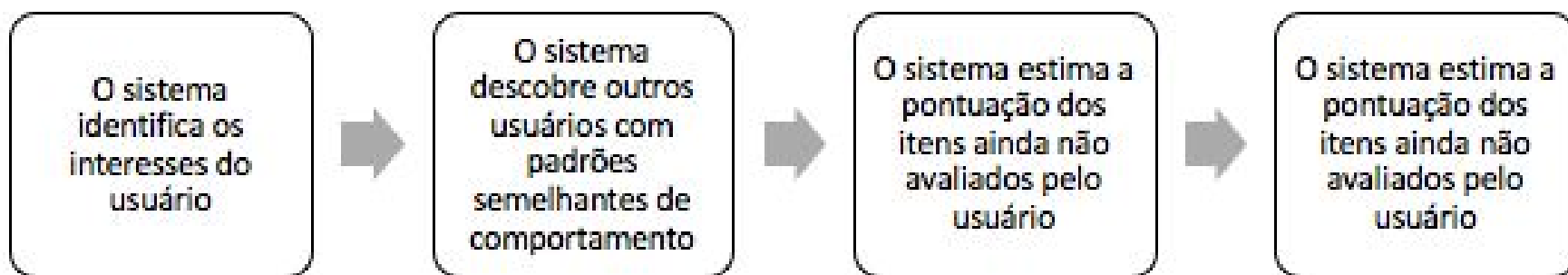
## Como funciona

- Não considera o conteúdo dos itens;
- Considera não apenas as preferências do usuário;
- A essência está na troca de experiências entre usuários com interesses em comum;
- Usuários semelhantes tendem a gostar das mesmas coisas!

# Como funciona



# Como funciona



# Tipos

- A fim de estabelecer recomendações, os sistemas CF precisam relacionar duas fundamentalmente diferentes entidades: itens e usuários.
- Há duas abordagens principais para facilitar essa comparação, que constituem as duas principais técnicas de CF: a abordagem de vizinhança e a abordagem de modelos de fatores latentes.

# Tipos

- Métodos de vizinhança focalizam as relações entre os itens ou alternativamente, entre os usuários.
- Uma abordagem item-item modela a preferência de um usuário para um item com base em classificações de itens semelhantes pelo mesmo usuário.
- Elevado custo computacional.
- É o mais utilizado.



# Tipos

- Modelos de fatores latentes, como a fatoração de matriz (aka, SVD), compreendem uma abordagem alternativa, transformando ambos os itens e usuários para o mesmo espaço fator latente.
- O espaço latente tenta explicar classificações caracterizando ambos os produtos e usuários sobre fatores inferidos automaticamente a partir de feedback do usuário.
- Modelo é usado para calcular previsões.
- Mais complexo e demorado.

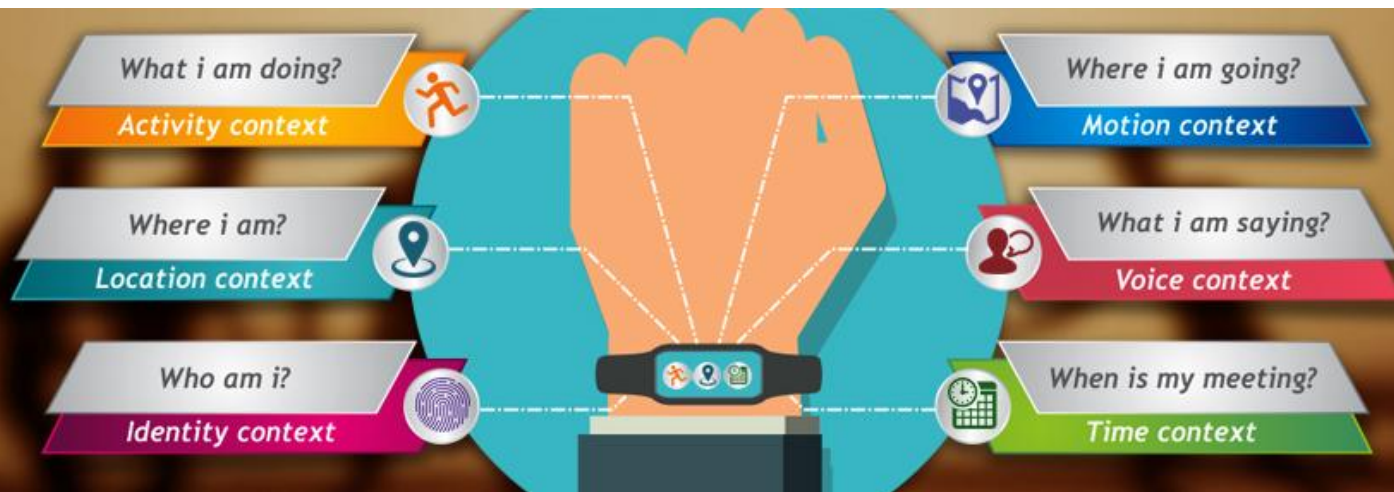
# Vantagens

- O sistema pode apresentar aos usuários recomendações inesperadas;
- Formação de comunidades de usuários.

# Limitações

- Um novo item nunca é recomendado até que um usuário o avalie;
- Um usuário exótico terá dificuldades para encontrar outros usuários com gostos similares;
- Necessita de uma vasta base de dados. Caso o número de usuários seja pequeno em relação ao número de itens, causará problema.

# SISTEMAS DE RECOMENDAÇÃO BASEADOS EM CONTEXTO



# Introdução

- A maioria das abordagens existentes para Sistemas de Recomendação se concentram em recomendar os itens mais relevantes a usuários individuais e não levam em consideração qualquer informação contextual, tais como tempo, lugar e da companhia de outras pessoas (por exemplo, para assistir filmes ou jantar fora).
- Em outras palavras, os sistemas de recomendação tradicionalmente lidam com aplicações que tem apenas dois tipos de entidades, usuários e itens, e não colocá-los em um contexto ao fornecer recomendações.

# O que é Contexto?

**O contexto é um conceito multifacetado, que tem sido estudado em diferentes linhas de pesquisa, incluindo informática (principalmente em inteligência artificial e computação ubíqua), ciência cognitiva, linguística, filosofia, psicologia e ciências organizacionais**

# Modelando Contexto

- Sistemas de Recomendação Sensíveis ao Contexto (CARS) lidam com a modelagem e previsão de gostos e preferências do usuário através da incorporação de informação contextual disponível para o processo de recomendação como explícitas categorias adicionais de dados.
- Estas preferências e gostos de longo prazo são geralmente expressas como classificações e são modeladas como a função não apenas de itens e usuários, mas também do contexto.

Usuário x Item x Contexto -> Recomendação

# Modelando Contexto: Exemplo

- Considerar a aplicação para recomendar filmes para usuários, onde os usuários e os filmes são descritos como as relações com os seguintes atributos:
- Filme; Usuário.



## Modelando Contexto: Exemplo

- Além disso, a informação contextual consiste nos três tipos seguintes, que também são definidas como as relações com os seguintes atributos:
- Teatro; Tempo; Companhia.
- Em seguida, o rating atribuído a um filme por uma pessoa também depende de onde e como o filme foi visto, com quem e em que momento

# Obtendo Contexto

A informação de contexto pode ser obtida num número de maneiras, incluindo:

- Explicitamente, isto é, ao abordar diretamente as pessoas relevantes e outras fontes de informação contextual e reunindo explicitamente esta informação, quer por meio de perguntas diretas ou provocando esta informação através de outros meios.
  - Por exemplo, um site pode obter informações contextuais, pedindo uma pessoa para preencher um formulário web ou para responder a algumas perguntas específicas antes de fornecer acesso a determinadas páginas da web.

# Obtendo Contexto

- Implicitamente a partir dos dados ou para o ambiente, como uma mudança na localização do usuário detectado por uma empresa de telefonia móvel. Alternativamente, a informação de contexto temporal, podem ser implicitamente obtido a partir da data e hora de uma transação.
  - Nada há a ser feito, nestes casos, em termos de interação com o utilizador ou outras fontes de informação contextual.
- Inferir o contexto usando métodos de mineração de dados ou estatística.

# Paradigmas para incorporar contexto em SR

- Diferentes abordagens para utilizar informações contextuais no processo de recomendação podem ser classificados em dois grupos:
  - (1) recomendação via consulta e pesquisa dirigida ao contexto, e
  - (2) recomendação via levantamento e estimativa de preferência contextual.

## Paradigmas para incorporar contexto em SR

- A abordagem de consulta e pesquisa orientada para o contexto tem sido usado por uma grande variedade de sistemas de recomendação móveis e turísticas.
- Os sistemas que utilizam esta abordagem normalmente usam informações contextuais (**obtido diretamente do usuário, por exemplo, a obtenção de hora local, tempo ou localização atual**) para consultar ou procurar um determinado repositório de recursos (**por exemplo, restaurantes**) e apresentar os melhores recursos correspondentes (**por exemplo, restaurantes que estão actualmente abertos**) para o usuário.

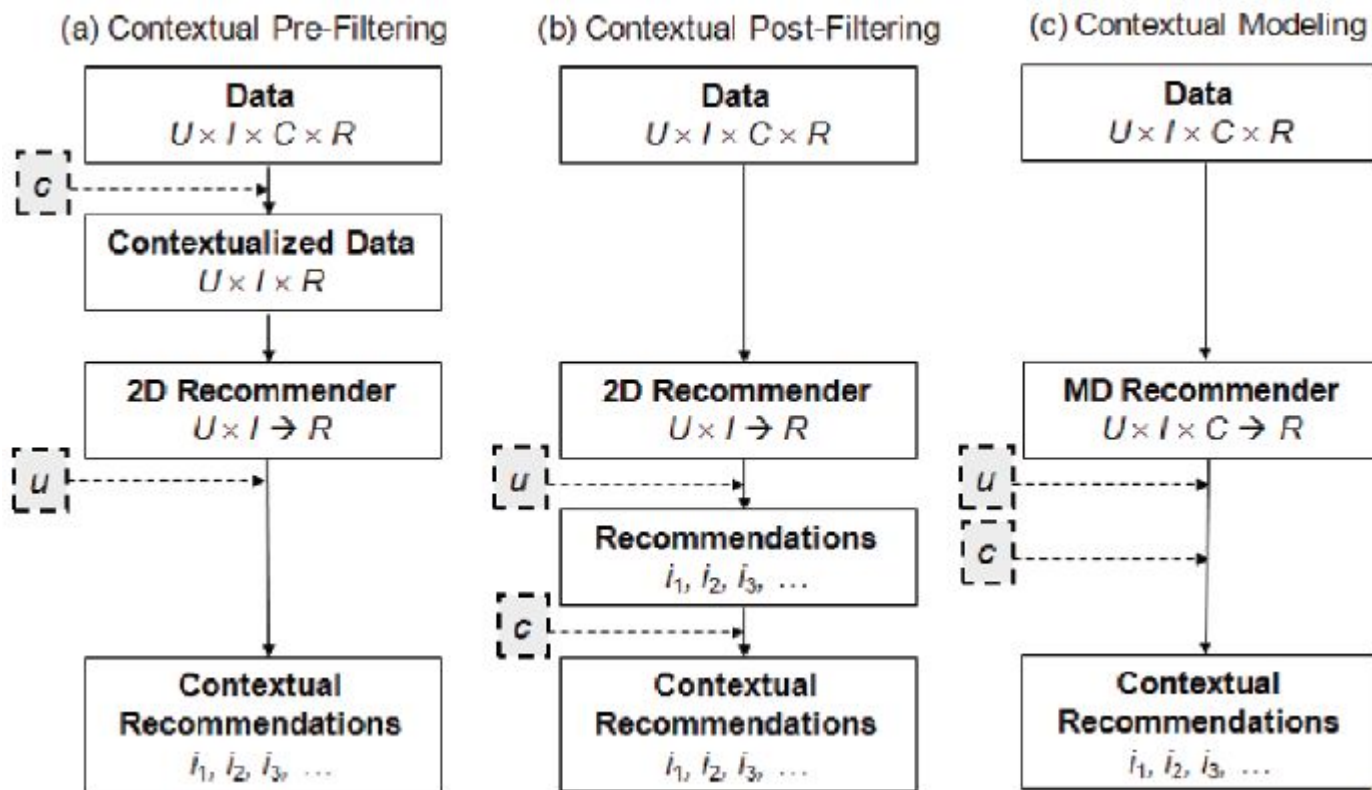
# Paradigmas para incorporar contexto em SR

- A abordagem de consulta e pesquisa orientada para o contexto pode ser feita das seguintes formas:
  - Pré-filtragem contextual: Neste paradigma recomendação, informações contextuais dirige a seleção de dados ou construção dos dados para esse contexto específico
  - Pós-filtragem contextual: Neste paradigma recomendação a informação contextual é inicialmente ignorada, e as classificações são previstas usando qualquer sistema 2D de recomendação tradicional sobre a totalidade dos dados.

# Paradigmas para incorporar contexto em SR

- A abordagem de consulta e pesquisa orientada para o contexto pode ser feita das seguintes formas:
  - Modelagem contextual: Neste paradigma a recomendação da informação de contexto é usado diretamente na técnica de modelagem como parte de estimativa de classificação.

# Paradigmas para incorporar contexto em SR





## Paradigmas para incorporar contexto em SR

- A outra abordagem: através de preferências levantamento e estimativa contextual, representa uma tendência mais recente na literatura de sistemas de recomendação sensíveis ao contexto.
- Para modelar as preferências sensíveis ao contexto dos usuários e gerar recomendações, estas técnicas tipicamente adotam filtragem colaborativa existente, baseado em conteúdo, ou métodos recomendação híbrido para configurações de recomendação sensíveis ao contexto ou aplicar várias técnicas de análise de dados inteligentes de mineração de dados ou aprendizagem de máquina.

# Combinando abordagens

- Muitas vezes uma combinação (a "mistura" ou um conjunto) de várias soluções oferece melhorias significativas de desempenho sobre as abordagens individuais.
- Os três paradigmas para sistemas de recomendação sensíveis ao contexto oferecem várias oportunidades para o emprego de abordagens combinadas.
  - Uma possibilidade é o de desenvolver e combinar vários modelos de um mesmo tipo.

# Combinando abordagens

- Os três paradigmas para sistemas de recomendação sensíveis ao contexto oferecem várias oportunidades para o emprego de abordagens combinadas.
  - Outra possibilidade interessante deriva de uma observação de que a informação contextual complexa pode ser dividida em vários componentes, e a utilidade de cada pedaço de informação contextual pode ser diferente dependendo se ele é usado na pré-filtragem, pós-filtragem, ou etapa de modelagem.

# AVALIAÇÃO DE DESEMPENHO



## Avaliação SR

- Avaliação: métricas, medidas, frameworks.
- Medidas, validações padrões e bases publicas.

# Propriedades SR

- Preferencias do usuário.
- Taxa de predição;
- Cobertura;
- Confiança;
- Robustez;
- Adaptatividade;
- Escalabilidade;
- Risco;
- Dentre outros...

# Tipos de Avaliação

- Medidas Qualitativas
- Medidas Probabilísticas
- Medidas de Ranking
- Usuários
- Outras

## Medidas Qualitativas

- Usada quando se quer minimizar número de erros.
- Diferentes tipos de sistemas de recomendações, preditores, recuperação de informação, etc.
- Itens são classificados como relevantes ou irrelevantes e uma métrica adotada é aplicada.
  - Acurácia, F-measure, Cobertura, diversidade, dentre outros.
- O problema de recomendação é tratado como tarefa de classificação.



# Medidas Probabilísticas

- Usadas especialmente para avaliar a confiança das previsões.
  - Ex.: probabilidade de ter recomendado item não relevante.
- O problema de recomendação é tratado como uma tarefa de regressão.
- Root Mean Square Error(RMSE), Mean Absolute Error(MAE)

# Medidas de Ranking

- O quão bem itens são ranqueados.
  - Precision and Recall Curves, Normalized Discounted Cumulative Gain(NDCG), Mean Average Precision(MAP), Hit Rate(HR), Fall Out, Area under the ROC Curve(AUC), etc.
- Avaliação da qualidade do rank dos itens retornados.
- Problema de recomendação tratado como tarefa de rank.

$$precision = \frac{1}{\#U} \sum_{u \in U} \frac{\#\{i \in Z_u | r_{u,i} \geq \theta\}}{n}$$

$$recall = \frac{1}{\#U} \sum_{u \in U} \frac{\#\{i \in Z_u | r_{u,i} \geq \theta\}}{\#\{i \in Z_u | r_{u,i} \geq \theta\} + \#\{i \in Z_u^c | r_{u,i} \geq \theta\}}$$

$$F1 = \frac{2 \times precision \times recall}{precision + recall}$$

# Usuários

- Experimentos empíricos com usuários para verificar satisfação deles quanto SR.
  - Experimentos offline , Estudos de Usuário, Validação online,
- Problemas: viés, falta de subjetividade, dificuldade de comparação de diferentes sistemas.

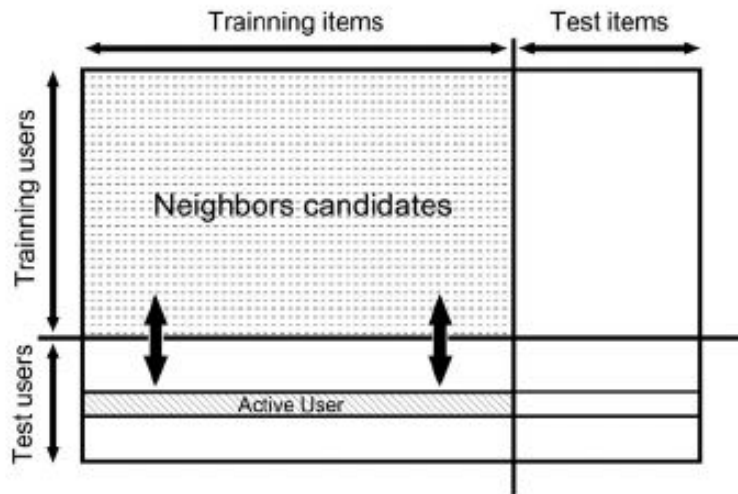
# Outros

- Frameworks
  - Pesos quanto a similaridade, significância, variância, etc.
  - Falta de formalização.
- *Diversity and novelty*
  - *Grau de diferenciação entre itens e grau de diferença entre itens recomendados e conhecidos pelo usuários.*
- *Entre outros...*

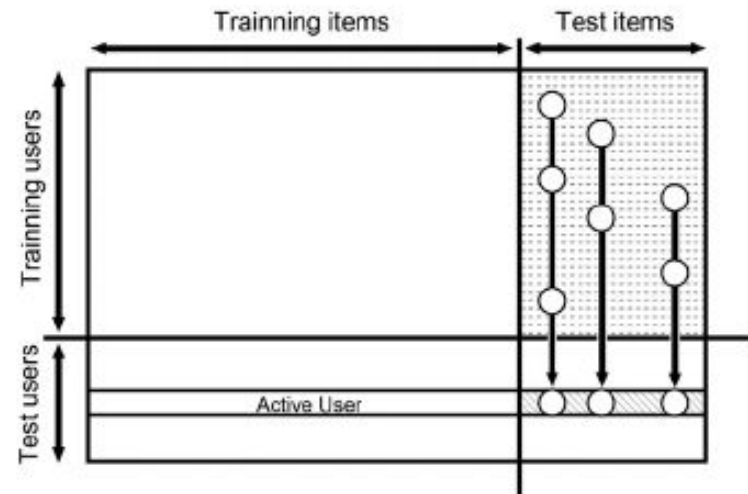
# Técnicas de Avaliação

- Usuário e Ranking mais usadas.
- Cada Algoritmo requer uma técnica apropriada.
- *Cross validation techniques*
- Importância!

### K-neighbors obtention



### Aggregation phase



### Evaluation metrics

<p><b>Predictions measures</b></p> <p>MAE = <math>(0.1 + 0.5 + 0.1) / 3 = 0.23</math></p> <p>RMSE = <math>\text{SQRT}((0.1^2 + 0.5^2 + 0.1^2) / 3) = 0.3</math></p> <p>Coverage = <math>100 \cdot 4 / 8 = 50.0 \%</math></p>
<p><b>Recommendation measures</b></p> <p>Precision = <math>2 / 3 = 0.66</math></p> <p>Recall = <math>2 / 4 = 0.50</math></p> <p>F1 = <math>(2 \cdot 0.66 \cdot 0.50) / (0.66 + 0.50) = 0.569</math></p>

### Error determination

4.9	1.5		4.1	3.2	KNN Predictions
5	1	3	4		Active user ratings

# SISTEMAS DE RECOMENDAÇÃO EM REDES SOCIAIS



# INTRODUÇÃO: WEB 2.0 E REDES SOCIAIS

- Web 2.0:
  - Tudo relacionado a pessoas
    - Criação de comunidades virtuais
    - Conectar pessoas às outras através de redes sociais.
    - Criação de conteúdos como wikis e blogs
    - Conteúdos com tags, comentários e avaliações
    - Permite que os usuários compartilhem conteúdo e interajam
- Redes Sociais
  - Design centralizado no usuário
  - Conteúdos gerados pelo usuário

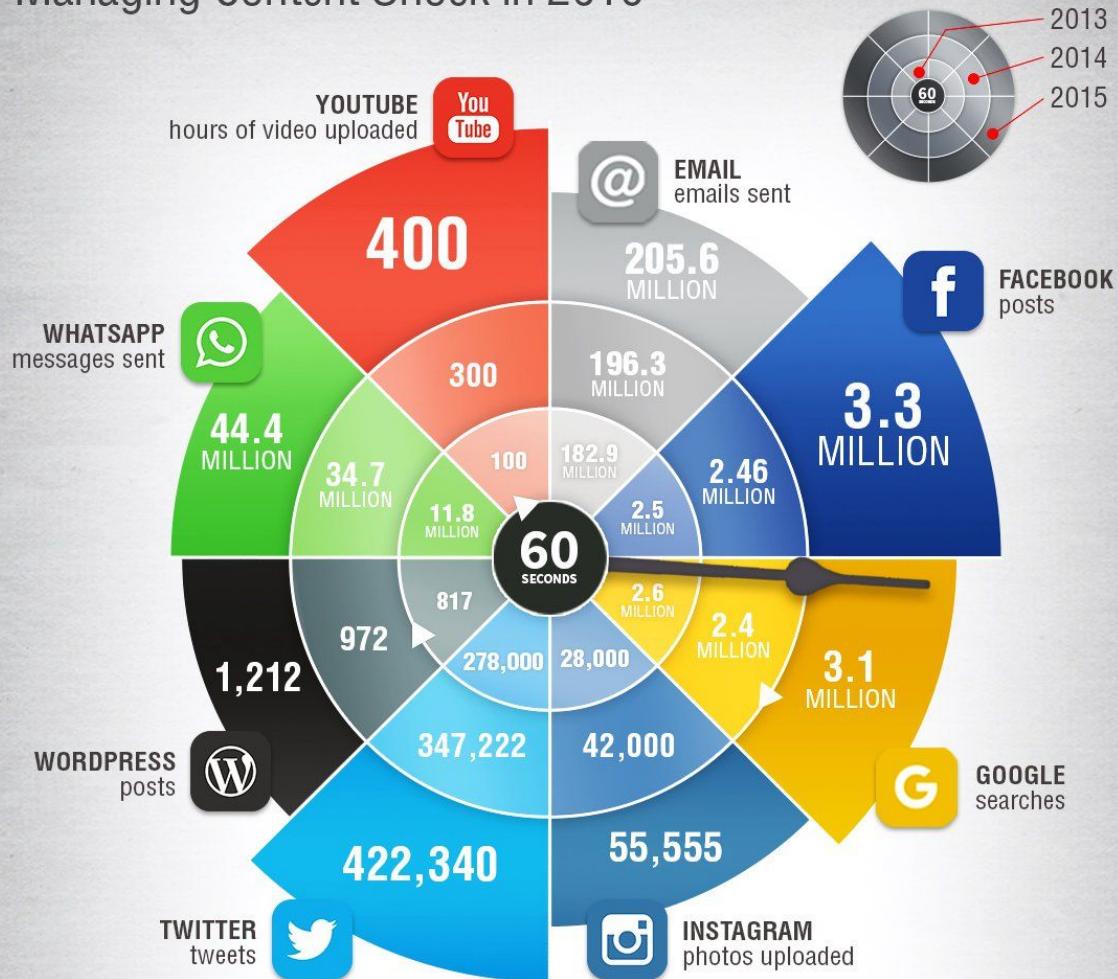


# INTRODUÇÃO: SITES MAIS ACESSADOS NO MUNDO

- 1 - Google.com
- 2 - Youtube.com
- 3 - Facebook.com
- 5 - Wikipedia.org
- 8 - Twitter.com
- 14 - LinkedIn.com
- 16 - Instagram.com
- 37- Stackoverflow.com

# What Happens Online in 60 Seconds?

Managing Content Shock in 2016





# INTRODUÇÃO: RECOMENDAÇÃO SOCIAL

- Um dos primeiros sistemas de recomendação sociais apareceu em 1997 <sup>1</sup>.
- Milhares de serviços de mídia social como o **Facebook** e o **Twitter** têm surgido nos últimos anos para permitir que as pessoas se comuniquem e se expressem.
- O rápido desenvolvimento dos meios de comunicação social tem acelerado muito o desenvolvimento de sistemas de recomendação social <sup>2</sup>

<sup>1</sup> Kautz, H., Selman, B., Shah, M.: Referral web: combining social networks and collaborative filtering. Communications of the ACM 40(3), 63–65 (1997) 76

<sup>2</sup> King, I., Lyu, M.R., Ma, H.: Introduction to social recommendation. In: Proceedings of the 19th international conference on World wide web, pp. 1355–1356. ACM (2010)

# INTRODUÇÃO: SISTEMAS DE RECOMENDAÇÃO SOCIAIS

- A rede social introduz **novos tipos de dados e metadados** que podem ser aproveitados pelo sistemas de recomendação (**marcações, comentários, curtidas, relações sociais explícitas**)
- Sistemas de recomendação podem afetar significativamente o sucesso de mídia social, **assegurando que cada usuário está servido com os itens mais relevantes** que se adaptem às suas necessidades pessoais.

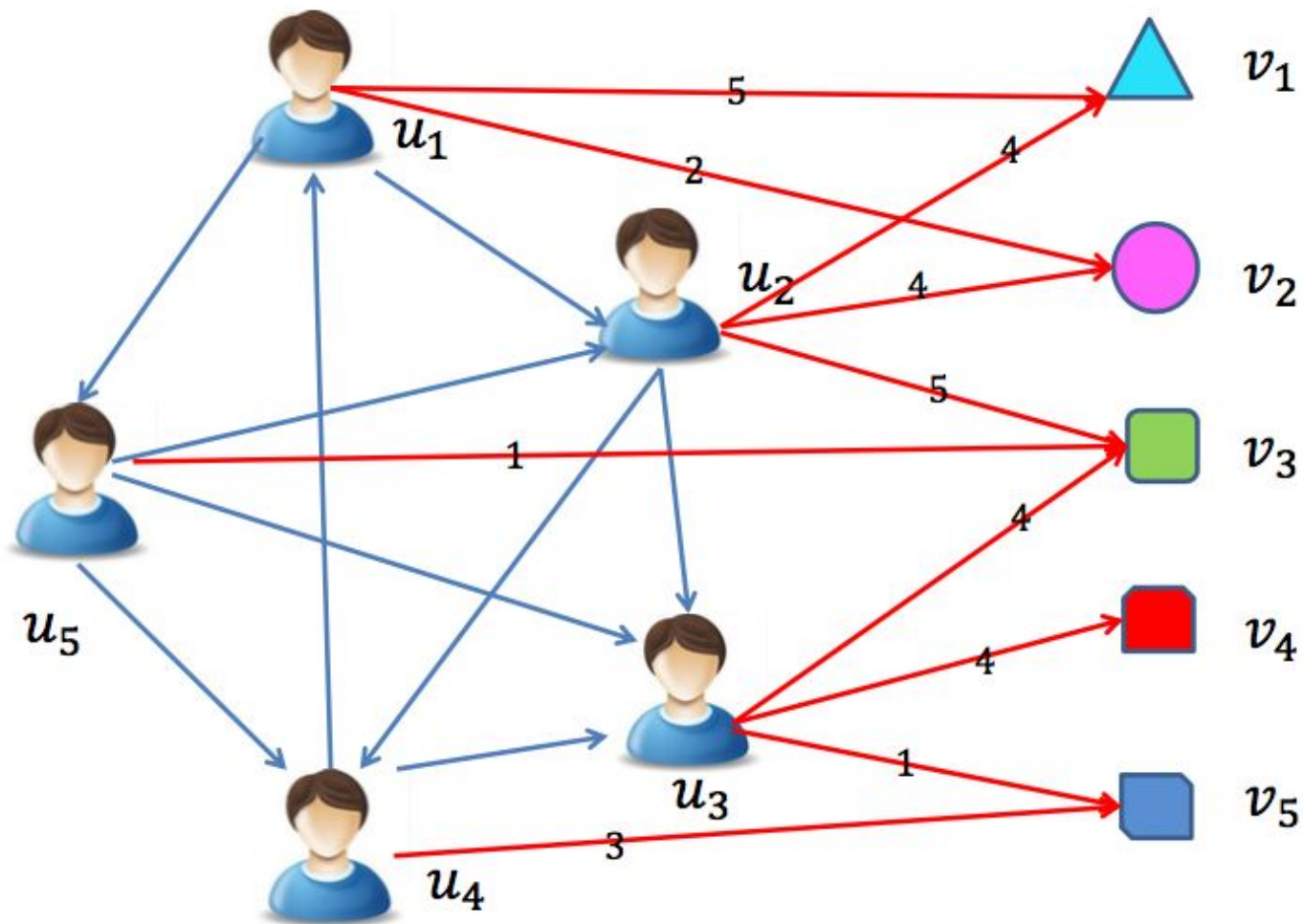
# DEFINIÇÕES

- **Qualquer recomendação que recebe as relações sociais como entrada adicional**, melhorando o mecanismo de recomendação existente com sinais sociais adicionais <sup>1</sup>.
- Relações sociais podem ser de <sup>1</sup>:
  - Confiança
  - Amizade
  - Membros de uma comunidade
  - Seguidores

# DEFINIÇÕES

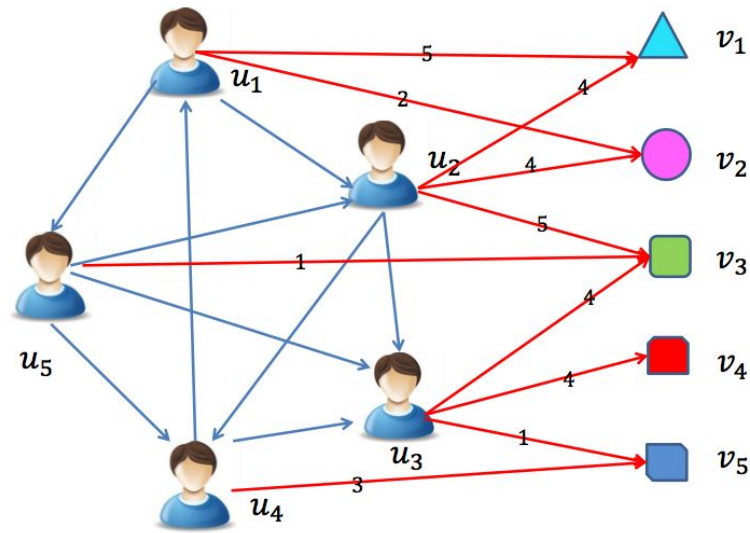
- **Qualquer sistema de recomendação que recomenda objetos do domínio social** <sup>1</sup>. Ou seja:
  - Tags
  - Pessoas
  - Comunidades

# DEFINIÇÕES



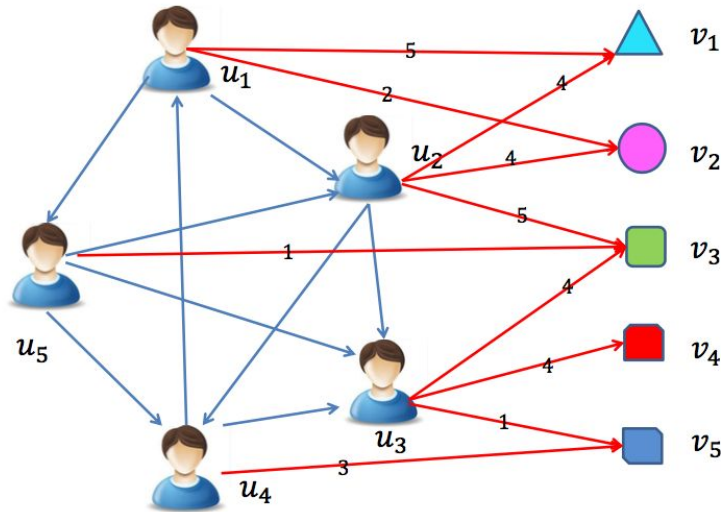


# DEFINIÇÕES: SISTEMAS DE RECOMENDAÇÃO TRADICIONAL



	$v_1$	$v_2$	$v_3$	$v_4$	$v_5$
$u_1$	5	?	2	?	?
$u_2$	4	4	5	?	?
$u_3$	?	?	4	4	1
$u_4$	?	?	?	?	3
$u_5$	?	?	1	?	?

# INTRODUÇÃO: SISTEMAS DE RECOMENDAÇÃO SOCIAIS



	$v_1$	$v_2$	$v_3$	$v_4$	$v_5$		$u_1$	$u_2$	$u_3$	$u_4$	$u_5$
$u_1$	5	?	2	?	?	$u_1$	0	1	0	0	1
$u_2$	4	4	5	?	?	$u_2$	0	0	1	1	0
$u_3$	?	?	4	4	1	$u_3$	0	0	0	0	0
$u_4$	?	?	?	?	3	$u_4$	1	0	1	0	0
$u_5$	?	?	1	?	?	$u_5$	0	1	1	1	0

# CONFIANÇA

- Confiança passou a ser nos últimos anos a ser um objeto de interesse de diversas áreas.
- Confiança é caracterizada quando as ações de um determinado usuário **influencia** na ação de outros usuários <sup>1</sup>.

# CONFIANÇA

- A confiança pode ser expressa de diversas formas e em diferentes níveis. De acordo com suas propriedades ela pode possuir as seguintes características:
  - **Subjetiva:** cada usuário pode possuir níveis de confiança diferentes sobre o mesmo usuário.
  - **Assimétrica:** o nível de confiança entre dois usuários não é necessariamente igual. Ou seja, um usuário A pode confiar mais em um usuário B do que B em A.
  - **Contextualizada:** quando a confiança é medida de acordo com um determinado contexto e não pode ser estendida diretamente para outros contextos.
  - **Transitiva:** quando se utiliza a confiança de um amigo confiável em um terceiro usuário, que não seja amigo do usuário em questão, para se definir a confiança.

# CONFIANÇA

- O cálculo do nível de confiança pode utilizar métricas globais e locais
  - **Local:** quando se utiliza métricas locais são analisadas apenas as opiniões subjetivas dos usuários. O grau de confiança local é diretamente proporcional a quantidade de indivíduos utilizados no cálculo. Custo computacional alto, mas com menos chance de erro
  - **Global:** analisam como uma determinada comunidade confia em um usuário. O custo computacional para processar a rede é menor por acontecer basicamente uma vez, mas **falhas** nesse processo podem ser constantes por existir pessoas não confiáveis que podem levar vantagem nessa métrica

# CONFIANÇA

- A predição de um nível de confiança entre usuários sem utilizar informações explícitas pode ser uma tarefa de grande complexidade. O fato de um usuário seguir ou compartilhar alguma postagem de outro não quer dizer necessariamente que existe confiança de um para com o outro.
- Além disso, a afirmação de confiança entre usuários produz muito mais impacto em sistemas de recomendação do que uma simples relação social <sup>1</sup>.

<sup>1</sup> YUAN, Q.; CHEN, L.; ZHAO, S. Factorization vs. regularization: fusing heterogeneous social relationships in top-n recommendation. In: ACM CONFERENCE ON RECOMMENDER SYSTEMS

# DESCONFIANÇA

- Muitas vezes ignorada por alguns trabalhos, deve ser considerada em modelos que utilizam confiança e principalmente quando fazem sua propagação.
- Alguns autores defendem que a desconfiança é apenas o oposto de confiança <sup>1</sup> e outros que ela pode ser a falta de confiança <sup>2</sup>.

<sup>1</sup> GANS, G. et al. Continuous requirements management for organisation networks: a (dis) trust-based approach. Requirements Engineering, [S.I.], v.8, n.1, p.4–22, 2003.

<sup>2</sup> STEPHEN, M. Formalising trust as a computational concept. Ph. n dissertation. University of Stirling, scotland, [S.I.], 1994.

# RECOMENDAÇÃO DE TAGS

- A adição de tags para objetos feita por usuários fornece um contexto adicional e informação semântica a vários recursos
- Tags são úteis para muitas aplicações:
  - Buscas, classificação, expansão de consulta
- Recomendar tags apropriadas a serem aplicadas pelo usuário para um determinado item pode:
  - Auxiliar o usuário na fase de etiquetagem
  - Reduzir o ruído indesejado



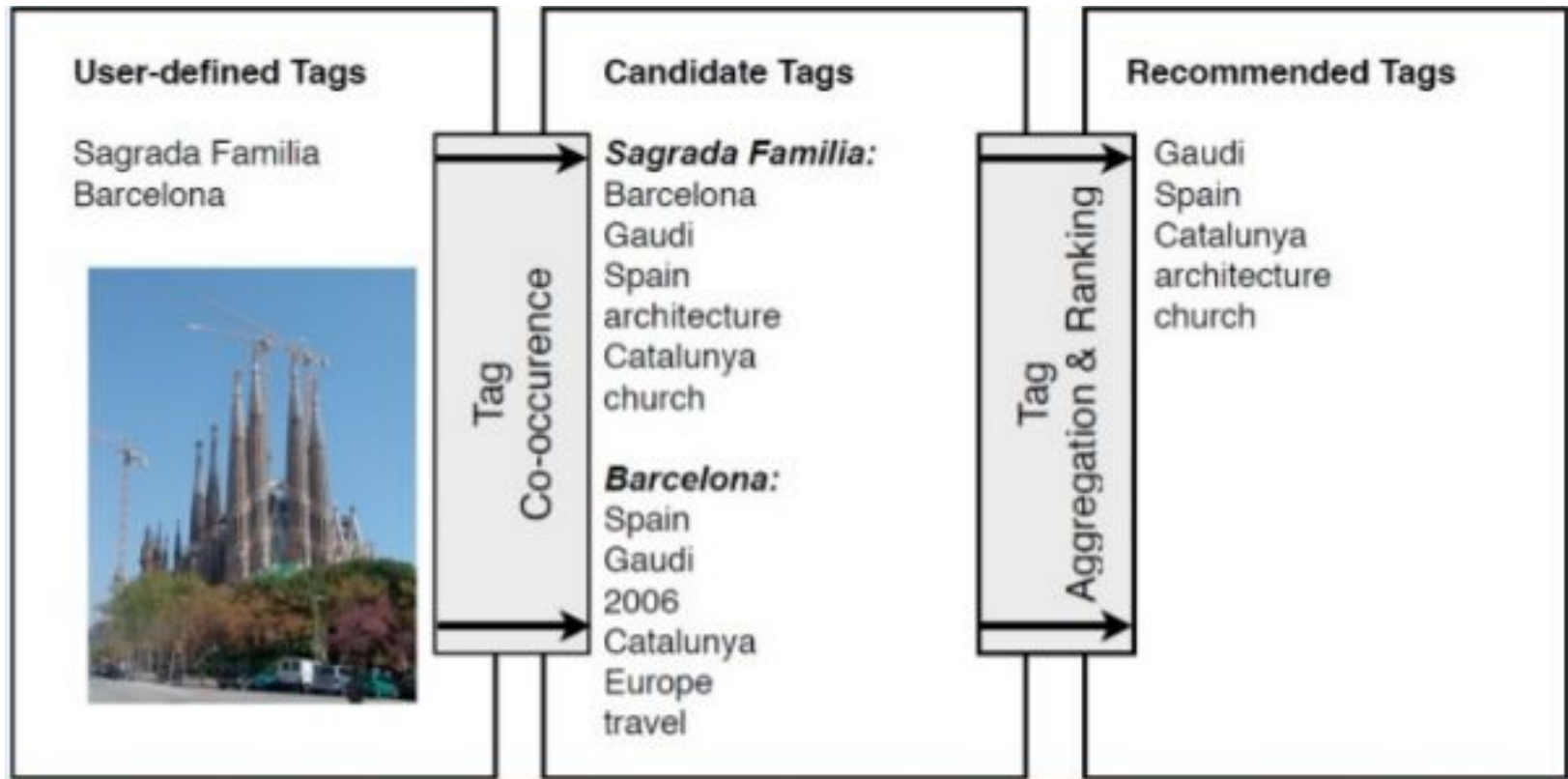
# RECOMENDAÇÃO DE TAGS - ESTRATÉGIAS

- Popular:
  - Recomenda-se etiquetas mais populares para o usuário:
    - Tags populares já atribuídas para o item alvo
    - Etiquetas frequentes anteriormente utilizado pelo utilizador
    - Etiquetas relacionadas com tags já atribuídos.
- Filtragem colaborativa:
  - Recomenda-se etiquetas associadas a itens "semelhantes"
  - Recomenda-se etiquetas dadas por usuários "semelhantes"
- Híbrido: recomenda-se etiquetas dadas pelos usuários semelhantes aos itens semelhantes

# RECOMENDAÇÃO DE TAGS - ESTRATÉGIAS

- Baseada em conteúdo:
  - Recomendar palavras-chave/frases do texto ou do item associado (conteúdo, metadados, etc.)
  - Analisar relação mútua entre conteúdo e tags:
    - Recomendar tags que têm a maior co-ocorrência com palavras-chave importantes

# FLICKR TAG RECOMMENDATION BASED ON COLLECTIVE KNOWLEDGE



# SISTEMAS DE RECOMENDAÇÃO SOCIAL EXISTENTES

- Geralmente, utilizam técnicas de filtragem colaborativa, especialmente as orientadas por usuário (user-based)
- A predição de avaliação é feita com auxílio de usuários correlatos feita em dois passos:
  - Primeiro: obtenção dos usuários correlacionados com um usuário alvo
  - Segundo: Filtragem colaborativa utilizando as avaliações dos usuários correlatos obtidas no primeiro passo
- Diferentes abordagens utilizam estratégias diferentes para obter os usuários correlatos

# EXPERIÊNCIAS **POSITIVAS** COM RECOMENDAÇÕES SOCIAIS

- Informações sociais contém informações complementares e os resultados de análise de redes sociais fornece o apoio técnico necessário para a recomendação social.
- Opiniões e gostos dos usuários podem ser propagadas através de redes sociais, o que pode reduzir o número de usuários com poucas avaliações ( cold-start ).
- Portanto, a recomendação social pode melhorar significativamente a cobertura da recomendação.

# EXPERIÊNCIAS **NEGATIVAS** COM RECOMENDAÇÕES SOCIAIS

- As relações sociais podem conter muito ruído e pode ter um impacto negativo sobre sistemas de recomendação;
- Usuários com poucas avaliações podem possuir poucos amigos também. Informações sociais podem não ser tão influentes na recomendação
- Diferentes tipos de relações sociais têm efeitos diferentes sobre sistemas de recomendação sociais e experiências de sucesso na exploração de relações de confiança não pode ser aplicável a outros tipos de relações.

# BASES DE DADOS E MÉTRICAS

Algorithms	Datasets							Metrics					Items		Users	
	D <sub>1</sub>	D <sub>2</sub>	D <sub>3</sub>	D <sub>4</sub>	D <sub>5</sub>	D <sub>6</sub>	D <sub>7</sub>	M <sub>1</sub>	M <sub>2</sub>	M <sub>3</sub>	M <sub>4</sub>	M <sub>5</sub>	I <sub>1</sub>	I <sub>2</sub>	U <sub>1</sub>	U <sub>2</sub>
SBWM[117,118]	✓							✓	✓	✓			✓			✓
TidalTrust[34]						✓						✓		✓		✓
MoleTrust[73-75]	✓								✓	✓				✓	✓	✓
TrustWalker[49]					✓			✓		✓	✓			✓	✓	✓
SoRec[67]	✓								✓					✓	✓	✓
LOCABAL[109]			✓	✓				✓	✓					✓		✓
STE[65]	✓							✓	✓					✓	✓	✓
mTrust[110]			✓	✓												
SocialMF[50]		✓			✓			✓						✓		✓
SoReg[68]	✓						✓	✓	✓					✓		✓

{D1(Epinions03), D2(Flixster), D3(Ciao), D4(Epinions11), D5(Epinions02), D6(FilmTrust), D7(Douban)}, {M1 (RMSE), M2 (MAE), M3 (Coverage), M4 (F-Measure), M5 (Others)}, {I1(Controversial Items), I2(Other Items)}, and {U1(Cold-start Users), U2(Other Users)}.

# RECUPERAÇÃO DE INFORMAÇÃO E SISTEMAS DE RECOMENDAÇÃO





# Recuperação de Informação e Sistemas de Recomendação

- Recuperação de Informação (RI) tem o objetivo de ajudar os usuários a armazenar e pesquisar várias formas de conteúdo, como textos, imagens e vídeos.
- Sistemas de recomendação e RI lidam com problemas de filtragem e classificação ( Ranking ).
- RI geralmente se concentra em desenvolver técnicas de recuperação globais, muitas vezes negligenciando as necessidades individuais e preferências dos utilizadores.

## Filtragem Baseada em Conteúdo == RI ??

- Filtragem baseada em conteúdo é um modelo que prediz a avaliação do usuário baseado nos dados do item.
  - Quanto um novo item é apresentado, o sistema verifica se o usuário vai gostar ou não daquele item baseado nos atributos do mesmo.
- Muitos sistemas desse tipo utilizam os atributos dos itens para realizar consultas
  - Tal método é inspirado na RI, cada item pode ser considerado um documento e documentos similares são recuperados.

# Recuperação de Informação e Sistemas de Recomendação

- Sistemas de recomendação aproveitam técnicas de de RI. ( i.e. Filtragem baseada em conteúdo ).
- RI lida com grandes repositórios de dados não estruturados de uma grande variedade de domínios
- Sistemas de recomendação lida com pequenos repositórios e geralmente de um domínio específico.
- RI lida com a localização de conteúdos relevantes e o usuário consegue avaliar a relevância dos conteúdos retornados
- Em Sistemas de Recomendação, o usuário não possui conhecimento suficiente para avaliar a relevância. Os<sup>99</sup> itens são novos e "inesperados"

# Recuperação de Informação e Sistemas de Recomendação

- SR e RI suportam diferentes estágios no processo de busca/descobrimento
- Imagine buscar uma câmera fotográfica no google e acessar um site especializado em câmeras.

# Recuperação de Informação



Todas Notícias Maps Imagens Vídeos Mais Ferramentas de pesquisa

Aproximadamente 264.000.000 resultados (0,61 segundos)

## Santa Cruz Futebol Clube

Liga: Brasileiro Série A Posição: 20

Brasileirão Série A  
quarta-feira, 19 de outubro, 20:45  
Estádio do Arruda



Santa Cruz

0 - 1  
ENC.

Botafogo



Todos os horários estão no: Horário de Brasília

Calendário e resultados

## Notícias em destaque



### Uillian Correia é pego no doping e está suspenso preventivamente

NE10 - 1 dia atrás

O volante Uillian Correia, do **Santa Cruz**, foi pego no exame antidoping e está suspenso de ...

Por causa de dívida de R\$ 350 mil com César Baiano, estádio do Santa Cruz vai a leilão de novo  
Supersport - 1 dia atrás

## Santa Cruz Futebol Clube



Santa Cruz Futebol Clube mais conhecido como Santa Cruz é uma agremiação poliesportiva brasileira, sediada no Recife. Fundada a 3 de fevereiro de 1914, é um dos mais tradicionais e populares clubes de futebol de Pernambuco e do Nordeste brasileiro. [Wikipédia](#)

**Técnico:** Adriano Teixeira

**Arena/Estádio:** Estádio do Arruda

**Fundada em:** 3 de fevereiro de 1914, Recife, Pernambuco

**Campeonatos:** Copa do Brasil de Futebol, Copa Sul-Americana, mais

## Lista de jogadores

Marcos da Silva França  
Atacante

11

# Sistemas de Recomendação



The image shows a screenshot of a Spotify playlist interface. At the top, there's a header with the word 'PLAYLIST' and the title 'Descobertas da Semana'. Below the title is a description: 'Sua mixtape semanal com novas músicas! Ouça algumas descobertas escolhidas especialmente para você. Atualizada toda segunda!'. It also says 'Criada por: Spotify • 30 músicas, 1 h 53 m'. There are three buttons: 'PLAY' (green), 'SEGUINDO' (white outline), and a three-dot menu icon. Below this is a search bar with 'Filtrar' and a 'Baixar' toggle switch. The main content is a table of songs with columns for 'MÚSICA', 'ARTISTA', and a date indicator.

MÚSICA	ARTISTA	
+ Gotas de Chuva	O Berço	2 dias atrás
+ Ex-Amor	Ney Matogrosso	2 dias atrás
+ Senhor cidadão	Tom Zé	2 dias atrás
+ Já Nem Sei Mais - Ao Vivo	Roupa Nova, Chitãozin...	2 dias atrás
+ Menina Carolina	Bebeto	2 dias atrás
+ Refém do Seu Amor	Diego Damasceno	2 dias atrás
+ Cheia de Manias - Ao Vivo	Gigantes do Samba	2 dias atrás

## Conclusão

- RI não é a mesma coisa que Sistemas de Recomendação.
- Ambos podem fornecer conteúdo personalizado que corresponda às suas necessidades
- A forma como você interage com eles que os distingue um do outro. Em RI, costuma-se chamar sistemas de recomendação de zero query



UNIVERSIDADE FEDERAL DE PERNAMBUCO  
CENTRO DE INFORMÁTICA  
IN1152 – RECUPERAÇÃO INTELIGENTE DE INFORMAÇÃO



# SISTEMAS DE RECOMENDAÇÃO

Alysson Bispo, Larissa Falcão, Micael Soares, Samuel Romeiro  
{abp2, lctf, msf3, scras}@cin.ufpe.br