

Universidade Federal de Minas Gerais  
Instituto de Ciências Exatas  
Departamento de Ciências da Computação

ANNA CHRISTINA DE CARVALHO GUIMARÃES

MONOGRAFIA DE PROJETO ORIENTADO EM COMPUTAÇÃO I

**CONSTRUINDO O MAPA DA TV DIGITAL**

Belo Horizonte – MG  
2012 / 1º semestre

Universidade Federal de Minas Gerais  
Instituto de Ciências Exatas  
Departamento de Ciências da Computação  
Curso de Bacharelado em Ciência da Computação

**CONSTRUINDO O MAPA DA TV DIGITAL**

por

Anna Christina de Carvalho Guimarães

Monografia de Projeto Orientado em Computação I

Apresentado como requisito da disciplina de Projeto Orientado em  
Computação I do Curso de Bacharelado em Ciência da Computação da UFMG

Prof. Dr. Olga N. Goussevskaia

Orientadora

Prof. Dr. Jussara Marques de Almeida

Co-Orientadora

Belo Horizonte – MG

2012 / 1º semestre

## **Agradecimentos**

Agradeço aos meus pais pela vida que me proporcionaram.

Aos meus professores e colegas, pela troca inestimável de conhecimentos.

“How do you know what you’re going to do till you do it?”

**J.D. Salinger, *The Catcher in the Rye***

## Resumo

Novas formas de transmissão de conteúdo da TV digital possibilitam o acesso a quantidades cada vez maiores de filmes e programas. Com a variedade crescente de opções de programação, cresce também a dificuldade de se organizar esses programas e de encontrar conteúdos de interesse. Informações sobre programas de TV existem em abundância, assim como informações sobre seu público, mas não há ainda um conjunto exaustivo de dados, nem um padrão que os governe. O objetivo deste trabalho foi dar início à construção desse conjunto, como base para estudos futuros sobre o domínio da TV digital.

Este trabalho descreve o processo de estudo, coleta e armazenamento de dados referentes a programas da TV digital e seus espectadores. A partir de fontes selecionadas da web, em particular, redes sociais voltadas ao compartilhamento de preferências e projetos abertos sobre entidades da TV, buscou-se obter o maior número possível de informações descritivas sobre os programas e sobre sua relação com usuários. O banco de dados obtido ao final desta etapa de trabalho cumpre seu objetivo de permitir a derivação de conclusões sobre a forma como programas de TV se relacionam entre si e com seus espectadores. Estas informações poderão ser utilizadas como base para a organização de diferentes programas de acordo com suas diferenças e semelhanças. Se organizados num espaço relacional, será, então, possível obter um mapa da TV digital.

**Palavras-chave:** *TV Digital, Sistemas de Recomendação, Filtros Colaborativos, Personalização de Conteúdo, Armazenamento de Dados, Bancos de Dados em Grafo*

## Abstract

Novel ways of broadcasting TV content allow for users to gain easier access to different movies and TV programs. With the ever-growing variety of program options, it also becomes increasingly difficult to sort through these programs and discover new and interesting content. There is a great deal of information about TV programs, as well as their spectators, but there is no all-encompassing data set, nor a standard they all fit into. Building such a data set was our goal with this project.

This paper focuses on the study, gathering and storage of data that refers to TV programs and their spectators. We focused our search for data on a few selective sources, particularly social networks that focus on sharing tastes and preferences in regards to entertainment and open projects about TV entities always aiming to retrieve as much descriptive information about different TV programs and their relationship with users. The database we were left with at the end of this first stage in our project serves its purpose and allows us to derive conclusions about how TV programs relate to one another and to their spectators. Such information can be used as a starting point for organizing programs according to their differences and similarities. If we were to sort them in a relational space, we would then hold a TV map.

**Keywords:** *Digital TV, Recommender Systems, Collaborative Filtering, Contents Personalization, Data Storage, Graph Database*

## Lista de Figuras

Figura 1	Modelo do Banco de Dados .....	17
Figura 2	Exemplo de um Banco de Dados da TV digital .....	18

## **Lista de Siglas**

API      Application Programming Interface

## Lista de Tabelas

Tabela 1	Comparação de desempenho de consultas entre um banco relacional e um banco em grafos, em função do número de relacionamentos percorridos e do tempo gasto para execução. ....	19
----------	---	----

# Sumário

<b>1</b>	<b>INTRODUÇÃO</b> .....	<b>9</b>
1.1	Visão geral .....	9
1.2	Justificativa e motivação .....	9
1.3	Objetivos e contribuições .....	10
<b>2</b>	<b>CONTEXTUALIZAÇÃO E TRABALHOS RELACIONADOS</b> .....	<b>11</b>
<b>3</b>	<b>METODOLOGIA</b> .....	<b>12</b>
3.1	Tipo de pesquisa .....	12
3.2	Procedimentos metodológicos .....	12
<b>4</b>	<b>DESENVOLVIMENTO</b> .....	<b>13</b>
4.1	Fontes e Coleta de Dados .....	13
4.2	Armazenamento de Dados .....	15
<b>5</b>	<b>RESULTADOS E DISCUSSÃO</b> .....	<b>19</b>
<b>6</b>	<b>CONCLUSÃO E TRABALHOS FUTUROS</b> .....	<b>21</b>
	<b>Referências</b> .....	<b>22</b>

# 1 INTRODUÇÃO

## 1.1 Visão geral

O surgimento de novas tecnologias de transmissão de vídeo e novos sistemas e serviços de TV digital implica a disponibilização de quantidades cada vez maiores de conteúdo aos seus usuários. Com uma variedade crescente de opções de canais e programas, torna-se difícil encontrar conteúdos de interesse. Em geral, guias em formatos impressos e digitais proveem alguma informação sobre a programação disponível, mas o usuário ainda precisa lidar com o problema de procurar programas específicos manualmente entre as muitas variedades existentes. Além disso, a falta de conhecimento prévio sobre cada programa, já que muitas vezes há apenas uma descrição breve dos mesmos, quando alguma, restringe o usuário a pequenos grupos de programas já conhecidos.

Até agora, sistemas de recomendação vêm sendo usados para contornar essa dificuldade, propondo aos usuários sugestões de programas que podem ser de seu interesse, mas frequentemente de forma restrita e pouco satisfatória. Além disso, pouco tem sido feito no sentido de construir uma visão geral da TV digital. A existência de uma coleção de dados sobre entidades da programação de TV seria, portanto, proveitosa para o desenvolvimento de soluções para a questão de sobrecarga dos usuários, além de poder servir como base para estudos sobre o relacionamento entre programas e suas propriedades, o que os aproxima e os distancia, e sobre o comportamento de usuários perante diferentes programas.

## 1.2 Justificativa e motivação

O problema de sobrecarga de conteúdo disponível ao usuário existe em outros domínios além da TV digital, já com algumas soluções que permitem atenuar seus efeitos negativos. Sites de comércio eletrônico, por exemplo, frequentemente armazenam o histórico de compras de seus clientes para fazer sugestões personalizadas de outros produtos, enquanto sites de distribuição de música e vídeo fazem recomendações com base em descrições textuais (*tags* e palavras-chave) da mídia acessada. Estas técnicas de recomendação permitem que usuários tenham acesso mais direto àquilo que é de seu interesse, ao invés de serem obrigados a percorrer listas de todo acervo de conteúdo.

A disponibilidade de informações existentes sobre programas de TV e seus espectadores leva à ideia de que seria possível aplicar técnicas semelhantes às usadas para recomendações de conteúdo para agrupar programas de acordo com sua similaridade. Apesar de já existirem sistemas de recomendação para programas de TV, estes ainda funcionam de forma isolada e por isso tem uso e sucesso limitados. Informações sobre programas e filmes encontram-se dispersas e podem ser altamente heterogêneas, enquanto informações sobre espectadores são restritas aos conjuntos de usuários de cada aplicação de distribuição ou recomendação de programas. Não há ainda uma base de dados completa que permita a derivação de conclusões sobre como programas e usuários se relacionam.

### **1.3 Objetivos e contribuições**

O objetivo deste trabalho é superar os limites e dificuldades até agora existentes para a organização de programas de TV e o estudo de seus relacionamentos, a partir da construção de um conjunto de dados que contenha não só informações específicas sobre programas, mas também sobre sua relação com usuários reais. Espera-se que esse conjunto de dados forneça uma visão global do mundo da TV digital, além de servir como base para aplicações futuras, desde sistemas de recomendação mais poderosos até classificadores automáticos.

## 2 CONTEXTUALIZAÇÃO E TRABALHOS RELACIONADOS

A idéia de organizar programas de televisão em um espaço relacional surgiu da experiência em [9], com a construção de um mapa da música a partir da identificação de coocorrências de pares de músicas em listas de preferências criadas por usuários e disponíveis na web. Estas listas de preferências mostram-se como fortes indicadores da correlação entre pares de produtos e vêm sendo usadas como base para problemas de recomendação em diferentes domínios, como os serviços da Amazon [12] e Netflix [19].

Além da análise de preferências de usuários com filtros colaborativos [1, 18, 17] para determinar o nível de correlação entre dois dados ou produtos, é possível também utilizar filtros baseados em conteúdo, que analisam descrições textuais dos dados e avaliam a similaridade entre os atributos de cada um [14], partindo do pressuposto que itens semelhantes serão descritos e categorizados de forma semelhante. A combinação dos dois tipos de filtros num filtro híbrido [6] tem-se mostrado vantajosa e vem sendo utilizada em um número de projetos com aplicações para TV digital, como visto em [5, 16, 3].

Em [8], propõe-se um guia de programação eletrônico pessoal, a partir da criação de perfis com as preferências de usuários e de descrições textuais dos programas assistidos. Mais relacionado à pesquisa aqui proposta é o sistema [queueo.tv](http://queueo.tv)<sup>1</sup>, descrito em [4]. Nesta aplicação, o usuário é contemplado diariamente com recomendações de programas geradas a partir do uso de filtros híbridos.

Entre as poucas aplicações existentes que vão além de sistemas de recomendação, pode-se destacar o aplicativo [Plizy](http://www.plizy.com)<sup>2</sup>, um agregador de vídeo que permite a criação, geração automática e compartilhamento de playlists de vídeos.

---

<sup>1</sup><http://queueo.tv>

<sup>2</sup><http://www.plizy.com>

## **3 METODOLOGIA**

### **3.1 Tipo de pesquisa**

De acordo com a classificação de [11], a pesquisa aqui proposta é de cunho aplicado, já que o resultado esperado é uma solução concreta do problema estudado. Com relação aos seus objetivos, a pesquisa pode ser considerada como exploratória, com o estudo do problema apresentado e de possíveis soluções a partir de bases preexistentes.

### **3.2 Procedimentos metodológicos**

Apesar da existência de um volume considerável de informações sobre filmes e programas, elas ainda estão muito dispersas e há poucos conjuntos de dados compreensivos sobre o domínio da TV digital. Assim, num primeiro momento de trabalho, foi necessário fazer um levantamento de possíveis fontes de dados. Nessa etapa, foram considerados fatores como ruído nas informações, volume, relevância e facilidade de obtenção dos dados. Após a identificação e seleção de fontes viáveis, iniciou-se a etapa de coleta propriamente dita, a partir da elaboração de *crawlers* (coletores de informações). Esta busca de dados será contínua durante todo o andamento do projeto, mesmo depois do início da implementação da rede de relações. À medida que novas informações forem obtidas, elas poderão ser adicionadas à base de dados já recuperados.

Foi preciso, também, estudar possibilidades de armazenamento para os dados, de forma a facilitar sua posterior manipulação. A grande quantidade de informações sobre programas da TV digital, juntamente com a falta de uma padronização para os mesmos e o relacionamento intrínseco entre eles e seus espectadores implica a necessidade de um banco de dados flexível e eficiente. Para este fim, investigou-se diferentes alternativas de armazenamento de dados em busca de uma que atendesse às demandas específicas do conjunto de dados a ser mantido e estudado. De alguns modelos promissores, decidiu-se desenvolver e implementar o modelo de banco de dados em grafos.

O desenvolvimento de cada uma das fases do trabalho é descrito em maiores detalhes na seção seguinte.

## 4 DESENVOLVIMENTO

A realização dos aspectos práticos do trabalho foi dividida em duas fases, uma que se concentrou na obtenção de dados e outra em que se construiu uma estrutura de armazenamento capaz de comportar estes dados.

### 4.1 Fontes e Coleta de Dados

Para a construção de um conjunto de dados referentes à programação da TV digital e seus espectadores, o primeiro passo foi identificar possíveis fontes para sua obtenção. Mais especificamente, buscou-se informações que tanto caracterizassem programas individualmente (atributos específicos, como título e elenco) quanto possibilitassem agrupar diferentes programas por similaridade (atributos compartilhados, como gênero, palavras-chave e país de origem). Visto que o comportamento de usuários em relação aos programas também é um indicador da semelhança entre eles, era preciso também obter informações sobre o que usuários de serviços da TV digital assistiam. Como esses dados não estão disponíveis diretamente, foi feito um levantamento de outras fontes que proovessem informações sobre os gostos e preferências de indivíduos com relação a filmes e programas de TV.

Entre diversas fontes avaliadas, decidiu-se restringir a coleta aos sites GetGlue<sup>1</sup>, TasteKid<sup>2</sup>, TMDb<sup>3</sup> e MovieLens<sup>4</sup>. A escolha por estes quatro domínios deveu-se principalmente à facilidade de obtenção de dados e à completude e organização dos mesmos.

#### 4.1.1 GetGlue

Nessa rede social, o usuário é capaz de visualizar diferentes listas de programas, separadas em categorias descritivas, e associar uma avaliação do tipo *Like* ou *Dislike* aos programas que ele já tenha assistido. Programas que receberam *Likes* podem ser subsequentemente adicionados a uma lista de programas favoritos criada pelo usuário.

---

<sup>1</sup><http://getglue.com>

<sup>2</sup><http://www.tastekid.com>

<sup>3</sup><http://www.tmdb.org>

<sup>4</sup><http://www.movielens.org>

Uma característica única da rede GetGlue é a possibilidade de fazer *Check-ins*, isto é, reportar quando se está assistindo um determinado filme ou programa. Os *Check-ins* são anunciados tanto nas páginas pessoais dos usuários, tornando-se visíveis para o próprio usuário e seus amigos na rede, quanto na página inicial do GetGlue, que mantém um *stream* de atualizações recentes dos usuários. *Check-ins* podem ser opcionalmente acompanhados de comentários, aos quais outros usuários podem responder interativamente.

Durante a coleta de dados para este trabalho, decidiu-se recuperar apenas as informações das listas de preferências dos usuários (contendo os programas classificados com *Like* ou *Dislike*). É possível, no entanto, que as atualizações de *Check-in*, em especial aquelas que acompanham algum comentário ou descrição, também compreendam informações úteis.

#### **4.1.2 TasteKid**

Este site é prioritariamente uma ferramenta de recomendação de filmes, programas, livros e música, porém também possui alguns aspectos que o fazem se assemelhar a uma rede social. Como ocorre no GetGlue, o usuário pode classificar programas e filmes que ele já tenha assistido, incluindo-o em listas de *Liked* e *Disliked*, de acordo com suas preferências. Estas listas também ficam visíveis na página do usuário. De forma inversa, é também possível acessar páginas de programas, ao invés de páginas de usuários. Nestas, pode-se ver uma pequena descrição do programa e uma lista de todos os usuários que adicionaram o programa às suas listas de preferências.

#### **4.1.3 The Movie Database**

TMDb é uma base de dados comunitária que agrupa informações detalhadas sobre filmes, incluindo descrições da trama, detalhes técnicos de produção, elenco e palavras-chave. Todo conteúdo é aberto para modificações e revisões por seus usuários. O site também possibilita a avaliação de filmes numa escala de 0.5 a 10 e mantém informações sobre a média das notas recebidas por cada filme, além das notas individuais recebidas de cada usuário.

#### 4.1.4 MovieLens

O MovieLens é um projeto desenvolvido especificamente para o estudo de algoritmos para sistemas de recomendação e filtragem de informação. Na prática, o site funciona como um sistema de recomendação que deriva sugestões de filmes a partir da aplicação de filtros colaborativos. Usuários que acessem o site podem avaliar filmes que eles tenha assistido, numa escala de 0.5 a 5, e essas notas são comparadas para encontrar outros grupos de filmes que tiveram notas semelhantes.

O grupo de pesquisa GroupLens, responsável pelo desenvolvimento do MovieLens, disponibiliza bases de dados coletados do site, abrangendo informações sobre títulos, notas de usuários e *tags*. Para este trabalho, utilizou-se uma das bases ofertadas pelo GroupLens, que continha mais de 10 milhões de avaliações de 10 mil filmes por 72 mil usuários. As informações sobre os filmes e usuários não são aprofundadas nesses conjuntos de dados, mas se mostram como um bom ponto de partida para este estudo, por descreverem a informação básica buscada: o relacionamento entre usuários e filmes. Os mesmos conjuntos de dados são utilizados também em outros trabalhos sobre sistemas de recomendações, como [13, 7].

Com exceção do MovieLens, as informações dos domínios citados anteriormente foram coletadas a partir de realização de chamadas dos métodos de suas respectivas APIs, disponibilizadas publicamente. Espera-se que a coleta de dados seja contínua ao longo de etapas futuras do trabalho, com novos dados sendo sempre adicionados à base já existente.

## 4.2 Armazenamento de Dados

Para armazenar o grande volume de informações que seria obtido ao longo da coleta, foi preciso construir um modelo de dados capaz de manter essas informações da forma mais simples e eficiente possível. Devido à natureza dos dados e à sua quantidade, assim como a forma como as entidades se relacionam umas com as outras, optou-se por abandonar o modelo relacional convencional de armazenamento de dados em favor da organização de dados em grafo, que ofereceria maior escalabilidade e performance, além de refletir melhor a natureza dos dados.

Um banco de dados em grafo modela informações em termos de nodos (instâncias de uma entidade ou objetos), relacionamentos, que ligam pares de objetos, e propriedades, que descrevem tanto nodos quanto relacionamentos. Como todos estes elementos são considerados igualmente importantes no grafo, esta modelagem é a mais intuitiva para representações

de dados altamente conectados e naturalmente estruturados. O banco de dados permite recuperar informações sobre relacionamentos de forma mais simples do que um banco de dados relacional, que requereria operações caras de junção de tabelas, e sua estrutura flexível permite que o modelo seja facilmente revisto para acomodar novas características [2].

### 4.2.1 Neo4j

Para a construção do banco utilizou-se o Neo4j<sup>5</sup>, um banco de dados NOSQL de alta performance que trabalha com uma estrutura de redes flexível orientada por objetos. Seu funcionamento se baseia em uma estrutura de grafos que armazena dados em nodos, arestas e propriedades, como explicado anteriormente. Essa estrutura é otimizada para a realizar caminhamentos no grafo (*graph traversals* [15]), permitindo recuperar informações sobre relacionamentos de forma eficiente e facilitando a implementação de algoritmos de grafos típicos, como o cálculo de caminho mínimo ou a centralidade de um nodo/entidade na rede. Casos de uso típico para o banco são aqueles que envolvem dados fortemente conectados, dados geográficos, redes sociais e algoritmos de recomendação.

Entre as vantagens do Neo4j que o separam de outros bancos de dados em grafos estão sua flexibilidade de implementação, que pode ser desenvolvida diretamente a partir da API Java ou a partir de *bindings* já existentes para outras linguagens, como Ruby e Python, a disponibilidade do código-fonte sob licença GPLv3, que facilita o estudo dos seus mecanismos internos de funcionamento, e a noção de transações [10] que delimitam as operações de modificação do banco. Estas últimas garantem a confiabilidade das operações no banco, em cumprimento das propriedades ACID de atomicidade, consistência, isolamento e durabilidade, e são pouco comuns em bancos de dados NOSQL, que priorizam disponibilidade e performance e abrem mão da consistência do banco a cada operação.

O banco de dados efetivamente construído durante este trabalho fez uso direto dos métodos do Neo4j disponíveis como bibliotecas para a linguagem Java e da linguagem Gremlin<sup>6</sup> de consulta e caminhamento em grafos, que permite expressar buscas profundas a partir de expressões simples.

---

<sup>5</sup><http://neo4j.org/>

<sup>6</sup><https://github.com/tinkerpop/gremlin/wiki>

## 4.2.2 Modelo de Dados

Para a base de dados deste trabalho, desejou-se armazenar informações relevantes sobre programas de TV e filmes, isto é, qualquer informação que pudesse ligar diferentes programas. Entre os atributos incluídos estão gênero/categoria, palavras-chave, tags, descrição e país de origem. À medida que mais dados foram obtidos, mais atributos poderão ser acrescentados ao grafo.

Também era necessário manter informações relativas aos espectadores e a forma como eles se conectam a diferentes programas. O fator mais importante aqui é o relacionamento entre usuários e entidades da TV, portanto não houve uma preocupação em guardar representações elaboradas dos usuários em si.

Uma simplificação do modelo de dados criado para este trabalho é apresentada na figura 1.

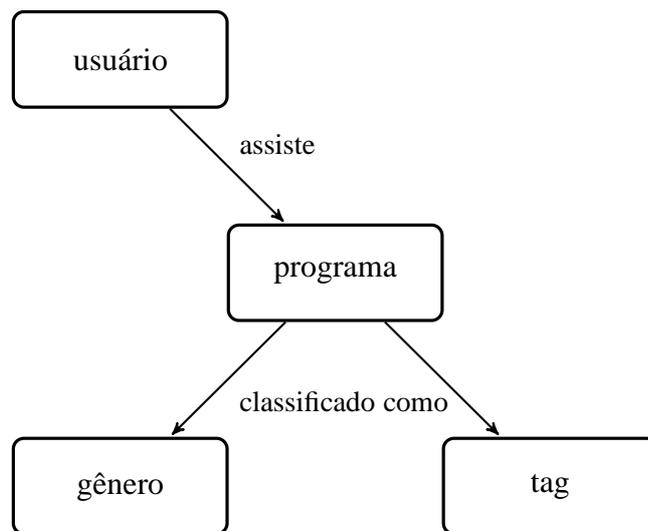


Figure 1: Modelo do Banco de Dados

Cada usuário presente no banco de dados é representado como um nodo, ou vértice do grafo, e seu relacionamento com os programas de TV por ele assistidos, também representados com nodos, é materializado na forma de uma aresta. Nodos representantes de programas também se ligam a nodos de descrições em formato de tags e gêneros, categorias estas comumente utilizadas para separar e organizar programas. Um exemplo de um pequeno banco de dados, compreendendo apenas um usuário, um filme e um programa, é mostrado na figura 2 para ilustrar a implementação do modelo.

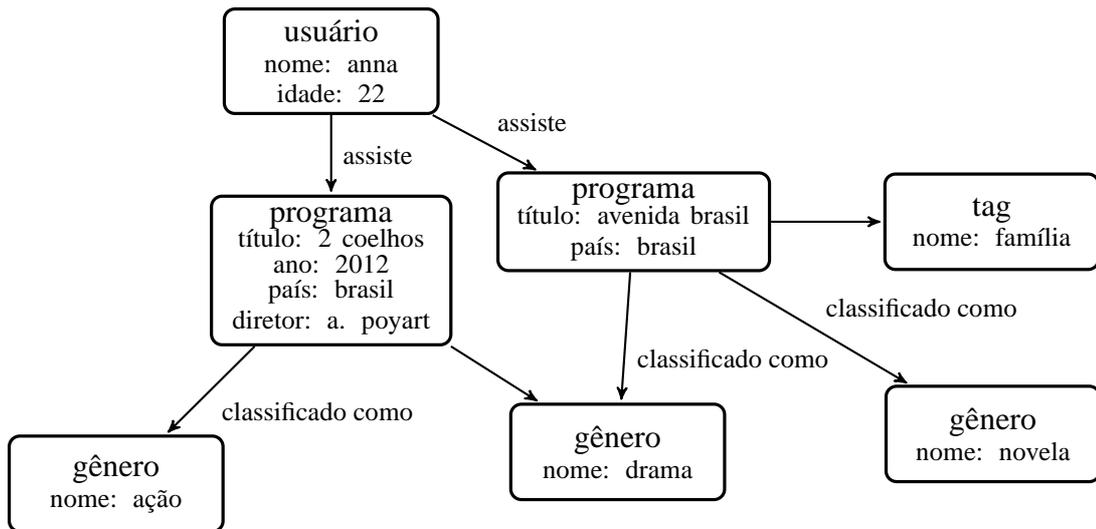


Figure 2: Exemplo de um Banco de Dados da TV digital

Até agora, o banco possui uma estrutura muito semelhante ao exemplo na figura 2, armazenando apenas informações essenciais sobre os dados - o suficiente para dar início a um estudo sobre o relacionamento entre programas de TV.

## 5 RESULTADOS E DISCUSSÃO

O resultado obtido com o desenvolvimento deste trabalho foi a implementação de um banco de dados, em pequena escala, de programas de TV. Apesar de as informações mantidas no banco ainda serem limitadas, tanto em quantidade quanto em nível de detalhes de cada entidade suportada, ele já permite a visualização da forma geral como usuários se relacionam com os diferentes programas aos quais eles tem acesso, e de como esses programas se relacionam entre si, a partir de características em comum.

Devido à forma como o banco de dados foi construído, o acréscimo de novos objetos e de novas propriedades e características aos dados já presentes no banco pode ser feito de forma simples e eficiente, sem a necessidade de refazer o esquema lógico que governa os dados. Assim, o modelo idealizado inicialmente pode ser facilmente estendido à medida que a coleta contínua de dados traga informações inéditas, ou que novas formas de relacionar os dados sejam descobertas.

Para verificar a escolha pelo armazenamento em um banco de dados em grafos, foram feitas comparações entre o banco desenvolvido e um banco relacional, ambos populados com dados sobre filmes e usuários obtidos do GroupLens. Procurou-se fazer consultas que se assemelhariam ao tipo de consulta real que será feita ao banco, para constatar a conformidade do banco às demandas específicas do projeto desenvolvido. Algumas das consultas são mostradas a seguir e seus resultados, em função do tempo de execução, são mostrados na tabela 1. Todos os testes foram realizados em condições semelhantes, no sistema operacional Ubuntu 10.04.2 LTS com processador Intel(R) Core(TM)2 Quad CPU Q9550 e 8GB de memória principal disponível.

Relacionamentos Percorridos	BD Relacional	BD em Grafos
17895	15.33s	0.1s
89475	0.28s	0.1s
16135	1760s	33s

Table 1: Comparação de desempenho de consultas entre um banco relacional e um banco em grafos, em função do número de relacionamentos percorridos e do tempo gasto para execução.

**Consulta 1.** Listar todos os usuários que assistem ao filme 'Toy Story'

**Consulta 2.** Listar todos os gêneros de filmes assistidos por usuários que assistem ao filme 'Toy Story'

**Consulta 3.** Calcular a distribuição de gêneros

Os testes realizados, ainda que para um conjunto de dados pequeno em relação ao tamanho final esperado do banco, já é capaz de demonstrar que a opção pelo armazenamento dos dados num modelo de grafos é vantajosa, principalmente quando são realizadas operações que lidam com as relações entre os dados. Quando os dados são tratados isoladamente, considerando-se apenas cada entidade num dado momento (por exemplo, buscar todos as datas de lançamento de todos os filmes) o desempenho de ambos os modelos é equivalente. No entanto, a forma explícita com que a estrutura de grafos mantém os relacionamentos dos dados leva um desempenho superior quando estas informações são requisitadas.

Além do ganho em tempo de execução, ressalta-se também que própria estrutura do banco permite explorar os dados de forma mais especializada, com maior eficiência. A partir de caminhamentos simples no grafo constituído pelo agrupamento das entidades de TV, é possível extrair conclusões sobre os dados que não seriam imediatamente aparentes. Em particular, o banco de dados permite encontrar coocorrências de programas e filmes entre aqueles assistidos por diferentes grupos de usuários, além da possibilitar a separação e o agrupamento de programas por suas propriedades e descrições. Estas informações são fortes indicadores da correlação que existe entre os dados e podem ser utilizadas como base para o estudo da similaridade entre programas e as razões por trás dessa similaridade, isto é, é possível descobrir quais programas são semelhantes a outros e por quê.

## 6 CONCLUSÃO E TRABALHOS FUTUROS

A base de dados iniciada neste trabalho constitui uma importante fonte de informações sobre entidades da TV digital, que até agora eram esparsas e pouco acessíveis. À medida que mais informações forem acrescentadas a essa base e o modelo dos dados for estendido para abranger um maior número de detalhes sobre os programas, espera-se que seja possível derivar conclusões mais especializadas sobre a forma como programas e usuários se relacionam. A coleta de novos dados e a evolução do modelo, que serão continuadas mesmo após o final desta etapa de trabalho, garantirão que a base iniciada caminhe para um conjunto de dados o mais completo possível.

A partir desta colação de dados, pretende-se iniciar um estudo mais aprofundado sobre a organização natural dos programas de TV por meio da sua disposição em um espaço relacional que demonstre os relacionamentos existentes entre diferentes títulos da programação. Especificamente, deseja-se testar se os métodos desenvolvidos em [9] seriam válidos também para o domínio da TV digital e como eles se comportariam diante de um conjunto distinto de dados.

A conclusão deste trabalho, portanto, abre caminho para diversos estudos que precisem ou desejem se basear em dados sobre programas de TV e seu público. Com o esclarecimento das relações existentes entre usuários, programas e suas características, será possível criar um mapa da TV digital que as retrate e permita uma visão geral e completa deste extenso domínio.

## Referências

- [1] ADOMAVICIUS, G., AND TUZHILIN, A. Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. *IEEE Trans. on Knowl. and Data Eng.* 17, 6 (June 2005), 734–749.
- [2] ANGLES, R., AND GUTIERREZ, C. Survey of graph database models. *ACM Comput. Surv.* 40, 1 (Feb. 2008), 1:1–1:39.
- [3] ARDISSONO, L., GENA, C., TORASSO, P., BELLIFEMINE, F., DIFINO, A., AND NEGRO, B. User modeling and recommendation techniques for personalized electronic program guides. In *Personalized Digital Television - Targeting Programs to Individual Viewers, volume 6 of Human-Computer Interaction Series, chapter 1* (2004), Kluwer Academic Publishers, pp. 3–26.
- [4] BARRAGÁNS-MARTÍNEZ, A. B., COSTA-MONTENEGRO, E., BURGUILLO, J. C., REYLÓPEZ, M., MIKIC-FONTE, F. A., AND PELETEIRO, A. A hybrid content-based and item-based collaborative filtering approach to recommend tv programs enhanced with singular value decomposition. *Inf. Sci.* 180, 22 (Nov. 2010), 4290–4311.
- [5] BUENO, D., CONEJO, R., MARTÍN, D., LEÓN, J., AND RECUENCO, J. G. What can i watch on tv tonight? In *Proceedings of the 5th international conference on Adaptive Hypermedia and Adaptive Web-Based Systems* (Berlin, Heidelberg, 2008), AH '08, Springer-Verlag, pp. 271–274.
- [6] BURKE, R. Hybrid recommender systems: Survey and experiments. *User Modeling and User-Adapted Interaction* 12, 4 (Nov. 2002), 331–370.
- [7] CHEN, Y., HARPER, F. M., KONSTAN, J., AND LI, S. X. Social comparisons and contributions to online communities: A field experiment on movielens. *American Economic Review* 100, 4 (2010), 1358–98.
- [8] EHRMANTRAUT, M., HÄRDER, T., WITTIG, H., AND STEINMETZ, R. The personal electronic program guide: towards the pre-selection of individual tv programs. In *Proceedings of the fifth international conference on Information and knowledge management* (New York, NY, USA, 1996), CIKM '96, ACM, pp. 243–250.
- [9] GOUSSEVSKAIA, O., KUHN, M., LORENZI, M., AND WATTENHOFER, R. From web to map: Exploring the world of music. In *Web Intelligence and Intelligent Agent Technology, 2008. WI-IAT '08. IEEE/WIC/ACM International Conference on* (dec. 2008), vol. 1, pp. 242–248.
- [10] GRAY, J., AND REUTER, A. *Transaction Processing: Concepts and Techniques*. Morgan Kaufmann Series in Data Management Systems. Morgan Kaufmann Publishers, 1993.
- [11] JUNG, C. F. *Metodologias para pesquisa e desenvolvimento: aplicadas a novas tecnologia, produtos e processo*. Axcel Books do Brasil Editora, Rio de Janeiro, 2004.

- [12] LINDEN, G., SMITH, B., AND YORK, J. Amazon.com recommendations: Item-to-item collaborative filtering. *IEEE Internet Computing* 7, 1 (Jan. 2003), 76–80.
- [13] MILLER, B. N., ALBERT, I., LAM, S. K., KONSTAN, J. A., AND RIEDL, J. MoviLens unplugged: experiences with an occasionally connected recommender system. In *Proceedings of the 8th international conference on Intelligent user interfaces* (New York, NY, USA, 2003), IUI '03, ACM, pp. 263–266.
- [14] PAZZANI, M. J., AND BILLSUS, D. Content-based recommendation systems. In *THE ADAPTIVE WEB: METHODS AND STRATEGIES OF WEB PERSONALIZATION. VOLUME 4321 OF LECTURE NOTES IN COMPUTER SCIENCE* (2007), Springer-Verlag, pp. 325–341.
- [15] RODRIGUEZ, M. A., AND NEUBAUER, P. The graph traversal pattern. *CoRR abs/1004.1001* (2010).
- [16] SALTER, J., AND ANTONOPOULOS, N. Cinemascreen recommender agent: Combining collaborative and content-based filtering. *IEEE Intelligent Systems* 21, 1 (Jan. 2006), 35–41.
- [17] SARWAR, B., KARYPIS, G., KONSTAN, J., AND RIEDL, J. Item-based collaborative filtering recommendation algorithms. In *Proceedings of the 10th international conference on World Wide Web* (New York, NY, USA, 2001), WWW '01, ACM, pp. 285–295.
- [18] SU, X., AND KHOSHGOFTAAR, T. M. A survey of collaborative filtering techniques. *Adv. in Artif. Intell.* 2009 (Jan. 2009), 4:2–4:2.
- [19] TÖSCHER, A., JAHRER, M., AND BELL, R. M. The bigchaos solution to the netflix grand prize, 2009.