Recomendação de Conteúdo em Ambientes de Convergência Digital Incorporada ao *Middleware* Ginga

Priscilla Kelly Machado Vieira
Laboratório de Aplicações de Vídeo Digital (LAVID)
Universidade Federal da Paraíba
Departamento de Informática
João Pessoa – Paraíba – Brazil
+55 (83) 3216-7093
priscilla@lavid.ufpb.br

Natasha Queiroz Lino
Laboratório de Aplicações de Vídeo Digital (LAVID)
Universidade Federal da Paraíba
Departamento de Informática
João Pessoa – Paraíba - Brazil
+55 (83) 3216-7093
natasha@ci.ufpb.br

ABSTRACT
The emerging scenario of interactive Digital TV (iDTV) is promoting the increase of interactivity in the communication process and also in audiovisual production, thus rising the number of channels and resources available to the user. This reality makes the task of finding the desired content becoming a costly and possibly ineffective action. The incorporation of recommender systems in the iDTV environment is emerging as a possible solution to this problem. This paper aims to propose an approach to content recommendation in iDTV, incorporated to the middleware Ginga, based on data mining clustering techniques and knowledge representation, considering the iDTV as a digital convergence environment (DTV and Web).

Categories and Subject Descriptors
H.3.5 [**Information Storage and Retrieval**]: On-line Information Services – data sharing, web-based services.

General Terms
Algorithms, Design.

Keywords
Interactive Digital TV, Ginga, Recommendation Systems.

1. INTRODUÇÃO
Com o advento da TV Digital interativa (TVDi), nota-se o aumento de interatividade no processo de comunicação além do incremento das produções audiovisuais [1], o que torna um ambiente favorável ao desenvolvimento de aplicações dedicadas a esta tecnologia. Neste sentido, a TV Digital (TVD) promoveu um cenário de aumento da quantidade de canais, serviços e conteúdo disponíveis ao usuário. Esta nova realidade faz com que a tarefa de encontrar o conteúdo desejado se torne uma ação onerosa e, em alguns casos, ineficiente. É neste contexto que Sistemas de Recomendação [2] emergem como solução possível para auxiliar esta escolha.

No contexto tratado, o trabalho propõe um módulo de recomendação adequado à arquitetura da TVDi brasileira, considerando-a como um ambiente de convergência digital (TV e Web) [3], inserido no *middleware* Ginga [4], permitindo o direcionamento de conteúdo de acordo com o usuário. O serviço de recomendação utilizará uma abordagem baseada em técnicas de Mineração de Dados [5] integradas a conceitos semânticos da Web Semântica [6].

2. CONTEXTO TEÓRICO
A evolução tecnológica associada com a inserção da Televisão Digital no Brasil propicia transformações profundas em todas as instâncias relativas ao meio, caracterizado por um processo de convergência a uma nova plataforma de comunicação baseada em tecnologias digitais de codificação, compressão, multiplexação, transporte, transmissão e recepção de informações. Adicionalmente, a TVDi congrega uma gama de possibilidades que incrementam o processo de comunicação: (i) Inclui a interatividade e (ii) Possibilita conectividade com a Web (TVDi Conectada). Estas características redemocratizam o acesso à informação [1], principalmente se considerarmos que a TV ainda é o principal meio de transmissão de informações e entretenimento para a população brasileira, estando presente em mais de 95% das residências; enquanto que os computadores possuem uma cobertura de pouco mais de 38% das residências brasileiras [7].

Arquiteturalmente, o sinal da TVD é recebido e decodificado por um dispositivo chamado de Set-Top-Box (STB). Por possuir tais funcionalidades, esse dispositivo assemelha-se a um computador com poder limitado de processamento, podendo, inclusive, ser conectado na Web. Torna-se permissivo, assim, o envio e recebimento de dados via STB. Considerando o Sistema Brasileiro de TV Digital (SBTVD) [8], uma das camadas que compõem a arquitetura do receptor é o *middleware*, responsável por abstrair as particularidades do sistema para aplicações e usuários. No cenário brasileiro, o software responsável por estas atividades é o Ginga [4].

Analisando o surgimento da TVD, observa-se o aumento da quantidade de canais e recursos disponíveis ao usuário [9]. Esta realidade propicia o desenvolvimento de sistemas de recomendação de conteúdo, que podem aperfeiçoar a busca por conteúdo, tornando-a menos onerosa e mais adequada aos interesses dos usuários.

2.1.1 Sistema de Recomendação de Conteúdo

Diversos aspectos devem ser considerados para a construção de um sistema de recomendação, a exemplo de como as informações são coletadas, e os métodos de recomendação empregados [10].

A coleta de dados implícita é transparente ao usuário, onde seu comportamento é monitorado por meio do histórico do sistema [11]. A principal vantagem desta abordagem é ser discreta ao objetivo do usuário, não sendo necessária sua interação direta com o sistema. Em contrapartida, a precisão do sistema é atenuada, pois depende intrinsecamente dos dados coletados ao longo do uso do sistema. Por outro lado, na coleta explícita o sistema necessita que o usuário informe claramente suas preferências [11]. A principal vantagem desta abordagem é a precisão para expressar preferências ou interesses. Em contrapartida, a necessidade de interação direta torna o processo muitas vezes dispendioso, sendo dificultado com a forma de entrada de dados via controle remoto. Esta abordagem pode não ser a mais adequada para o domínio da TV, por ser caracterizado, predominantemente, por interações passivas dos usuários [11].

Entre as técnicas utilizadas para recomendação de conteúdo inclui-se a Mineração de Dados [5], passo mais importante no processo de Descoberta de Conhecimento em Banco de Dados (em inglês, *Knowledge Discovery in Data Bases* - KDD) [12].

KDD pode ser definido como o processo não trivial de identificação de padrões válidos, potencialmente úteis em um conjunto de dados [12]. Esses padrões extraídos devem ser confiáveis, compreensíveis e úteis, podendo ser empregados nos âmbitos científico ou governamental.

A etapa de Mineração de Dados, inclusa no processo de KDD, consiste na aplicação de análise de dados e algoritmos que produz uma relação particular de padrões de dados [12]. Um das tarefas de mineração de dados é a Clusterização [13], que consiste em dado uma base de dados X, agrupar (clusterizar) os objetos de X de modo que os mais similares sejam alocados no mesmo grupo (cluster) e os menos similares em clusters distintos.

Observando a gama de dados a serem minerados e as informações que podem ser extraídas após a mineração, pode-se raciocinar automaticamente sobre estes dados estruturando-os em Ontologias [14].

Ontologias podem ser definidas como um modelo abstrato capaz de organizar de forma explícita e formal os conceitos e restrições relacionados a um domínio de interesse. Desta forma, por serem formais, Ontologias possibilitam o processamento automático por máquinas e facilitam o compartilhamento de informações por um grupo de pessoas ou máquinas [15].

Comumente os sistemas de recomendação são classificados de acordo com o modo que realiza o serviço. Classicamente, utilizam: (i) Filtragem Baseada em Conteúdo [16], parte do princípio de que, se um usuário visualizou determinado programa, é provável que no futuro goste de outros similares. (ii) Filtragem Colaborativa [16], parte do princípio de que grupos de pessoas similares possuem comportamentos semelhantes e (iii) Filtragem Hibrida [16], combinação de abordagens.

Mais recentemente, surgiram os sistemas de recomendação semântica, os quais são caracterizados por utilizarem uma abordagem semântica, que utiliza conceitos e técnicas de Representação de Conhecimento (RC) [17], como por exemplo, Ontologias, durante o processo de recomendação. RC é a subárea da Inteligência Artificial que avalia como o conhecimento pode ser representado simbolicamente e manipulado de forma automática por máquinas [17].

Em sistemas de recomendação, em geral, são utilizadas bases para armazenamento e gerenciamento das informações dos usuários [18, 19]. Adicionalmente, estas bases podem ser definidas por meio de ontologias, de forma a manter os dados padronizados, permitindo o compartilhamento de informações, provendo semântica e inferência automática de relacionamentos entre os dados [18].

3. IDENTIFICAÇÃO DO PROBLEMA

Esta proposta está inserida no projeto *Knowledge TV* (KTV) [3] que propõe uma camada semântica, baseada nos conceitos da Web Semântica, para prover serviços na plataforma da TVDi. Desta forma, dentro do contexto analisado, propomos a recomendação de conteúdo incorporada ao *middleware* Ginga. Este serviço deverá ser baseado em Clusterização de dados utilizando a técnica *K-means* [20], algoritmo de agrupamento caracterizado por ser popular, com desempenho linear e largamente utilizado em ambientes multimídia [18].

Todos os dados resultantes do processo de Clusterização deverão ser estruturados em uma ontologia, permitindo o compartilhamento de informações entre sistemas e a inferência sobre o conhecimento armazenado.

Para a definição do trabalho proposto foram considerados diversos requisitos, tais como: (i) Ser multi usuário, considerando a TV como um ambiente de vários usuários, (ii) Ser multi plataforma, podendo recomendar conteúdo tanto em outros sistemas de TV quanto em outros ambientes, como a web (iii) Ser genérico, possibilitando o serviço de recomendação independente do ambiente provedor dos dados.

Considerando os requisitos anteriores foi proposta a arquitetura baseada em módulos da Figura 1.

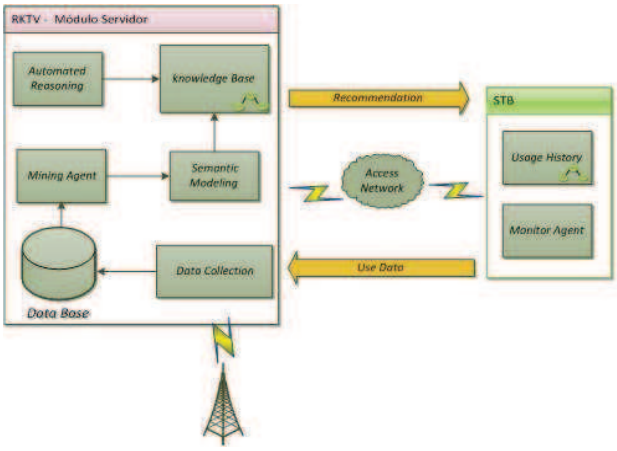

Figura 1. Arquitetura proposta.

A abordagem é baseada em uma arquitetura cliente-servidor. No lado servidor serão especificados todos os módulos que formarão o componente de recomendação proposto e inserido na arquitetura do KTV o *Recommender Knowledge TV* (RKTV). Desta forma, a arquitetura é estruturada em 7 (sete) módulos (Figura 1), distribuídos em 2 (dois) componentes: (i) o cliente, representado por um STB e (ii) o servidor, local onde será realizado o

processamento para a disponibilização do serviço de recomendação. Cada módulo possui funcionalidades especificas, como destacado a seguir:

- *Monitor Agent*: Monitorar o comportamento dos usuários do STB em relação ao conteúdo exibido na TVDi Conectada;
- *Usage History:* Responsável por estruturar e gerenciar em uma ontologia as informações da programação visualizada no STB (atividade realizada no Monitor Agent);
- *Data Collection:* Capturar os dados da programação das emissoras e dos usuários;
- *Mining Agent:* Responsável por todo o processo de mineração de dados sobre os dados coletados do STB e da emissora;
- *Semantic Modeling*: Tornar homogênea a descrição do conteúdo multimídia minerado, estruturado por meio de ontologias;
- *Knowledge Base*: Armazenar todo o conhecimento gerado no processo proposto em forma de ontologias;
- *Automated Reasoning*: Responsável por processar as informações presentes nas bases de conhecimento, gerando recomendações de conteúdo compatíveis aos interesses do usuário.

Os módulos do componente STB, apresentado na Figura 1, serão integrados ao Núcleo Comum do *middleware* Ginga onde são definidas funções básicas de sistemas de TVD, tais como: exibição e controle de mídias, controle de recursos do sistema, canal de retorno, dispositivos de armazenamento, acesso a informações de serviço e sintonização de canais [4]. A conexão entre o componente STB e o componente RKTV permitirá o serviço de recomendação semântica ao ambiente da TVD.

O módulo *Monitor Agent* atuará de forma implícita [11], monitorando o comportamento dos usuários do STB. Esta coleta de dados será por meio de uma comunicação constante com os módulos *Tuner* [4], responsável pela sintonização dos canais, e *SI* [4], responsável por extrair os metadados transmitidos pelas emissoras, do *middleware* Ginga.

O módulo *Data Collection* do componente RKTV captará a programação semanal das emissoras. Estes dados serão submetidos ao processo de Mineração de Dados. Após a atividade de Clusterização, grupos de conteúdo inter-relacionados serão gerados.

Os dados do histórico do uso do STB serão minerados, da mesma forma que os dados coletados da emissora. Duas formas de recomendação poderão ser geradas: (i) Os grupos de dados do histórico serão avaliados e qualquer conteúdo que estiver inserido em grupos semelhantes, gerados na mineração dos dados das emissoras, poderão ser recomendados, o que possibilitaria recomendações inesperadas e novas ao usuário. (ii) O conteúdo visualizado no momento em que a recomendação for solicitada será analisado e serão recomendados conteúdos que estejam classificados no mesmo grupo do conteúdo atual, o que possibilitaria recomendações mais próximas do esperado pelo o usuário. O raciocínio descrito será realizado sobre as ontologias, resultantes do processo de Mineração de Dados, no módulo *Knowledge Base* do componente RKTV. Basicamente, propomos um sistema de recomendação Hibrido, baseado em Filtragem Colaborativa e Semântica.

4. OBJETIVOS

Como objetivo geral propõe-se um sistema de recomendação adequado a arquitetura da TVDi brasileira, permitindo o direcionamento de conteúdo de acordo com o usuário. O trabalho inclui uma abordagem para recomendação em ambientes de convergência (TV e Web). Essa abordagem, baseada em conceitos de Representação do Conhecimento, como Ontologias, também propõe a integração de tecnologias da Web Semântica [21], permitindo, assim, o raciocínio automático sobre os dados e o compartilhamento de informação do sistema proposto com outros.

Para a obtenção do objetivo geral, foram definidos os seguintes objetivos específicos:

- Permitir recomendação de conteúdo multimídia em TVDi Conectada utilizando conceitos e técnicas da Web Semântica [6];
- Possibilitar que o *middleware* Ginga forneça o serviço de recomendação à TVDi Conectada;
- Obter um sistema de recomendação multi usuário, realizando recomendação de conteúdo multimídia a partir do uso do STB;
- Propor uma arquitetura genérica de recomendação de conteúdo multimídia, possibilitando o serviço independente do ambiente provedor dos dados.

5. CONTRIBUIÇÕE ESPERADAS

Ao final deste trabalho espera-se avançar no estado da arte em termos dos métodos para recomendação de conteúdo em TVDi, incorporando ao contexto da TV conceitos e técnicas baseadas nos princípios da Web Semântica. De modo suplementar, espera-se estender o *middleware* Ginga, com a incorporação de um módulo que disponha serviço de recomendação de conteúdo.

6. METODOLOGIA

A pesquisa deve consistir das seguintes etapas:

I. Realizar revisão bibliográfica relativa a métodos de recomendação no ambiente da TVD bem como à Descoberta de Conhecimento em Banco de Dados por meio de ontologias;

II. Propor um mecanismo de recomendação semântica, baseado em clusterização dos dados e ontologias, no ambiente da TVDi;

III. Desenvolver um módulo que possa ser inserido na arquitetura do SBTVD e que possibilite a recomendação de conteúdo;

IV. Avaliar a eficácia e relevância do módulo desenvolvido no item III.

7. ESTÁGIO DO TRABALHO

Atualmente, foi finalizado o levantamento do estado da arte em sistemas de recomendação de conteúdo. Foram avaliadas as técnicas e as características do processo de recomendação de trabalhos relacionados. A próxima etapa sob desenvolvimento trata da definição, implementação e avaliação do algoritmo de clusterização sobre os dados coletados do usuário e das emissoras de TV. Adicionalmente será definido o modelo de ontologia que estruturará o conhecimento proveniente da mineração, possibilitando, assim, o raciocínio automático sobre os dados.

Como trabalho futuro, destaca-se a necessidade de se implantar algoritmos de diversas tarefas da mineração de dados, como por

exemplo, classificação, regras de associação, entre outros, que permitirão um comparativo com a proposta, sendo possível, inclusive, a combinação de técnicas, fornecendo solução para problemas de múltiplas áreas. Também poderá ser estudada a integração desta solução em outros *middlewares* de sistemas de TVD, e em outros ambientes, tais como a web. Outro ponto que poderá ser explorado é o raciocínio avançado sobre a ontologia proposta neste trabalho, especificação de novos conceitos e relacionamentos que proporcionem melhor precisão no processo de recomendação.

8. TRABALHOS RELACIONADOS

Outros trabalhos propuseram sistemas de recomendação na TV. Ávila [9] apresenta um sistema integrado ao *middleware* Ginga baseado em mineração por regras de associação sem nenhuma estruturação semântica do conteúdo. Diferentemente, este trabalho propõe o uso de ontologias para estruturar o conhecimento após o processo de clusterização dos dados.

Aroyo [19] propõe a incorporação da semântica na TVD, considerando-a no ambiente da Web, a fim de possibilitar o desenvolvimento de aplicações distribuídas e a personalização do serviço. Aroyo baseia sua proposta de personalização nos conceitos da Web Semântica e na Mineração de Dados. Assim como Aroyo, nossa proposta também é baseada nos conceitos da Web semântica, no entanto, propomos recomendação no ambiente do SBTVD incorporando no *midleware* Ginga um módulo responsável por prover o serviço proposto.

9. REFERÊNCIAS

[1] Médola, A. S. L.. Televisão Digital Brasileira e os Novos Processos de Produção de Conteúdos: Os Desafios para o Comunicador. Revista da Associação Nacional dos Programas de Pós-Graduação em Comunicação | E-compós, Brasília, v.12, n.3, set./dez. 2009.

[2] Resnick, P. and Varian, H. R. 1997. Recommender systems. Communications of the ACM. ACM 40, 3 (Mar. 1997), 56-58.

[3] Lino, N., Araújo, J., Anabuki, D., JR, J. P., Batista, M., Nóbrega, R., Amaro, M., Siebra, C. and Lemos, G.. Knowledge TV. EuroITV, Lisboa, Portugal, 2011.

[4] Soares, L. F.; Lemos, G. Interactive Television in Brazil: System Software and the Digital Divide. In European Interactive TV Conference - EuroITV2007. Amsterdam, 2007.

[5] Han, J. and Kamber, M. Data Mining Concepts and Techniques, 2ª edição, Editora Elsevier, Reino Unido.

[6] Berners-Lee,T,; Lasila, O. Hendler, J. The Semantic Web. Scientific American, 284 (5):34-43, 2011.

[7] IBGE - Instituto Brasileiro de Geografia e Estatística. Síntese de Indicadores Sociais. Brasil, 2012.

[8] SBTVD – Sistema Brasileiro de TV Digital. Ministério das Comunicações. Disponível em: <http://sbtvd.cpqd.com.br/>. Acessado em 20 de Junho de 2012.

[9] Ávila, P. M. RecommenderTV: Suporte ao Desenvolvimento de Aplicações de Recomendação para o Sistema Brasileiro de TV Digital. Dissertaçao de Mestrado. Universidade Federal de São Carlos, Brasil, 2010.

[10] Ricc, F.; Rokach, L.; Shapira, B.; Kantor, B.P.. Recommender Systems Handbook. Springer Sciense Busines Media, LLC, 2011

[11] Chorianopoulos, K.. Personalized and mobile digital TV applications. Proc. 2008 MultimediaTools and Applications, Volume 36, Issue 1-2. Kluwer Academic Publishers, pp. 1-10, 2008.

[12] Fayyad, U. M., PIATESKY-SHAPIRO, G., SMYTH, P. From Data Mining to Knowledge Discovery: an Overview". Advances in Knowledge Discovery and Data Mining, AAAI Press, 1996.

[13] Ochi, L. S., DIAS, C. R., Stênio, S. F. Clusterização em Mineração de Dados. *Escola Regional de Informática Rio de Janeiro*. Espírito Santo. Mini Curso. Rio das Ostras, 2004.

[14] Guarino, N. Formal Ontology, Conceptual Analysis and Knowledge Representation, International Journal of Human-Computer Studies, 43(5-6):625–640, 1995.

[15] Gruber, Thomas R., A Translation Approach to Portable Ontology. Journal Knowledge Acquisition – Special issue: Current issues in knowledge modeling, 1993, v: 5. 199-220p.

[16] Adomavicius, G., Tuzhilin, A. Towards the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions. IEEE Transactions on Knowledge and Data Engineering, 2005, v: 17. 734-749p.

[17] Russell, S. and Norvig, P. Artificial Intelligence - A Modern Approach. 2 ª edição, Prentice Hall, 2002.

[18] Kim, J., Kwon, E., Cho, Y., Kang, S.. Recommendation System of IPTV TV Program Using Ontlogy and K-means Clustering. Ubiquitous Computing and Multimedia Applications, Daejeon, Korea, 2011.

[19] Aroyo, L.; Conconi, A.; Dietze, S.; Kaptein, A.; et *al*. NoTube – Making TV a medium for personalized interaction. EuroITV2009, Leuven, Belgium, 03-05 Jun 2009.

[20] Berkhin. P.. Survey of Clustering Data Mining Techniques. Technical report, Accrue Software, San Jose, CA, 2002.

[21] Breitman, K. . Web Semântica: O Futuro da Internet. 1. ed. Rio de Janeiro: LTC - Livros Tecnicos e Cientificos Editora S.A., 2005. v. 1. 190 p.